

# THE ITERATED CARMICHAEL LAMBDA FUNCTION

NICK HARLAND

**ABSTRACT.** The Carmichael lambda function  $\lambda(n)$  is defined to be the smallest positive integer  $m$  such that  $a^m$  is congruent to 1 modulo  $n$ , for all  $a$  and  $n$  relatively prime. The function  $\lambda_k(n)$  is defined to be the  $k$ th iterate of  $\lambda(n)$ . Previous results show a normal order for  $n/\lambda_k(n)$  where  $k = 1, 2$ . We will show a normal order for all  $k$ .

## 1. INTRODUCTION

The Carmichael lambda function  $\lambda(n)$  is defined to be the order of the largest cyclic subgroup of the multiplicative subgroup  $(\mathbb{Z}/n\mathbb{Z})^\times$ . It can be computed using the identity  $\lambda(\text{lcm}\{a, b\}) = \text{lcm}\{\lambda(a), \lambda(b)\}$  and its values at prime powers which are  $\lambda(p^k) = \phi(p^k) = p^k - p^{k-1}$  for odd primes  $p$  and  $\lambda(2) = 1, \lambda(4) = 2$ , and  $\lambda(2^k) = \phi(2^k)/2 = 2^{k-2}$  for  $k \geq 3$ .

Several properties of  $\lambda(n)$  were studied by Erdős, Pomerance, and Schmutz in [3]. In particular they showed that  $\lambda(n) = n \exp(-(1+o(1)) \log \log n \log \log \log n)$  as  $n \rightarrow \infty$  for almost all  $n$ . Martin and Pomerance showed in [6] that  $\lambda(\lambda(n)) = n \exp(-(1+o(1))(\log \log n)^2 \log \log \log n)$  as  $n \rightarrow \infty$  for almost all  $n$ . The  $k$ -fold iterated Carmichael lambda function is defined recursively to be

$$\lambda_1(n) = \lambda(n), \quad \lambda_k(n) = \lambda(\lambda_{k-1}(n)).$$

We define  $\phi_k(n)$  similarly. In [6] it is conjectured that

$$\lambda_k(n) = n \exp\left(-\frac{1}{(k-1)!}(1+o_k(1))(\log \log n)^k \log \log \log n\right)$$

for almost all  $n$ . In this paper we prove that conjecture.

**Theorem 1.** *For fixed  $k$ , the normal order of  $\log \frac{n}{\lambda_k(n)}$  is  $\frac{1}{(k-1)!}(\log \log n)^k \log \log \log n$ .*

We'll actually prove the theorem in the following slightly stronger form. Given any function  $\psi(x) = o(\log \log \log x)$  and  $\psi(x) \rightarrow \infty$  as  $x \rightarrow \infty$  we have

$$\log\left(\frac{n}{\lambda_k(n)}\right) = \frac{1}{(k-1)!}(\log \log n)^k \left(\log \log \log n + O_k(\psi(n))\right)$$

for all but  $O(x/\psi(x))$  integers up to  $x$ .

We will also turn our attention to finding an asymptotic formula involving iterates involving  $\lambda$  and  $\phi$ . Banks, Luca, Săidăk, and Stanic in [1] showed that for almost all  $n$ ,

$$\begin{aligned} \lambda(\phi(n)) &= n \exp(-(1+o(1))(\log \log n)^2 \log \log \log n) \text{ and} \\ \phi(\lambda(n)) &= n \exp(-(1+o(1))(\log \log n) \log \log \log n). \end{aligned}$$

As a corollary to Theorem 1 we will obtain asymptotic formulas for higher iterates involving  $\lambda$  and  $\phi$ . Specifically we prove the following.

---

2010 *Mathematics Subject Classification.* 11N56 (11N37).

**Theorem 2.** For  $l \geq 0$  and  $k \geq 1$ , let  $g(n) = \phi_l(\lambda(f(n)))$ , where  $f(n)$  is a  $(k-1)$  iterated arithmetic function consisting of iterates of  $\phi$  and  $\lambda$ . Then the normal order of  $\log(n/g(n))$  is  $\frac{1}{(k-1)!}(\log \log n)^k \log \log \log n$ .

An example of the use of this theorem is for  $\phi\phi\lambda\phi\phi\lambda\phi(n)$ . Since  $l = 2, k = 5$ , we get that the normal order of  $\log \frac{n}{\phi\phi\lambda\phi\phi\lambda\phi(n)}$  is

$$\frac{1}{4!}(\log \log n)^5 \log \log \log n.$$

The proof of Theorem 1 involves breaking down  $\frac{n}{\lambda_k(n)}$  in terms of the iterated Euler  $\phi$  function by using

$$(1) \quad \frac{n}{\lambda_k(n)} = \left( \frac{n}{\phi(n)} \right) \left( \frac{\phi(n)}{\phi_2(n)} \right) \cdots \left( \frac{\phi_{k-1}(n)}{\phi_k(n)} \right) \left( \frac{\phi_k(n)}{\lambda_k(n)} \right)$$

of which estimates for all but the last term are known. Hence  $\log \frac{n}{\lambda_k(n)}$  can be written as a sum of the logarithms on the right side of (1) and so we'll analyze the term  $\log(\phi_k(n)/\lambda_k(n))$ . The following notations and conventions will be used throughout the paper. The letters  $p, q, r$  will always denote primes and  $k \geq 2$  will be a fixed integer. Note that the theorem has already been proven for  $k = 1$ . Let  $v_p(n)$  be the largest power of  $p$  which divides  $n$ , so that

$$n = \prod_p p^{v_p(n)}.$$

Let the set  $\mathcal{P}_n$  be  $\{p : p \equiv 1 \pmod{n}\}$ . Throughout the paper we will assume  $x > e^{e^e}$  and  $y = y(x) = \log \log x$ . Also let  $\psi(x)$  be any function going to  $\infty$  such that  $\psi(x) = o(\log y) = o(\log \log \log x)$ . Whenever we use the phrase “for almost all  $n \leq x$ ” in a result, we mean that the result is true for all  $n \leq x$  except a set of size  $O(x/\psi(x))$ . Lastly we note that any implicit constant may depend on  $k$ .

## 2. REQUIRED ESTIMATES

The following estimates will be used throughout the paper. We use the Chebeshev bound

$$(2) \quad \sum_{n \leq x} \Lambda(n) = \sum_{p \leq x} \log p \ll x$$

where  $\Lambda(n)$  is the von-Mangoldt function. We also require a formula of Mertens (See [7, Theorem 2.7(b)])

$$(3) \quad \sum_{q \leq x} \frac{\log q}{q} = \log x + O(1).$$

Using partial summation on (2) we can obtain the tail estimates

$$(4) \quad \sum_{q > x} \frac{\log q}{q^2} \ll \frac{1}{x}$$

and

$$(5) \quad \sum_{q>x} \frac{1}{q^2} \ll \frac{1}{x \log x}.$$

Given  $m, x$ , let  $A$  be the smallest  $a$  for which  $m^a > x$ . We can then manipulate the sums

$$\sum_{a \in \mathbb{N}} \frac{P(a)}{m^a} = \frac{1}{m} \sum_{a=0}^{\infty} \frac{P(a)}{m^a} \text{ and } \sum_{\substack{a \in \mathbb{N} \\ m^a > x}} \frac{P(a)}{m^a} \ll \frac{1}{x} \left| \sum_{a=0}^{\infty} \frac{P(a)}{m^{a-A}} \right| = \frac{1}{x} \left| \sum_{a=A}^{\infty} \frac{Q(a)}{m^a} \right|$$

for  $Q(x) = P(x + A)$ . Then by noting that  $\sum_{a=A}^{\infty} \frac{P(a)}{m^a} \ll_P 1$  uniformly for  $m \geq 2$  and  $A \geq 0$  we obtain the estimates

$$(6) \quad \sum_{a \in \mathbb{N}} \frac{P(a)}{m^a} \ll_P \frac{1}{m}, \quad \sum_{\substack{a \in \mathbb{N} \\ m^a > x}} \frac{P(a)}{m^a} \ll_P \frac{1}{x}.$$

From [7, Corollary 1.15] we get

$$(7) \quad \sum_{s \leq x} \frac{1}{s} = \log x + O(1)$$

from which it easily follows that

$$(8) \quad \sum_{\substack{D \leq s \leq x \\ s \equiv a \pmod{C}}} \frac{1}{s} \ll \frac{1}{D} + \frac{\log x}{C}.$$

We will also make frequent use of the Brun-Titchmarsh inequality [7, Theorem 3.9]

$$(9) \quad \pi(t; n, a) \ll \frac{t}{\phi(n) \log(t/n)}.$$

By partial summation on (9) we can obtain

$$(10) \quad \sum_{\substack{p \leq t \\ p \in \mathcal{P}_n}} \frac{1}{p} \ll \frac{\log \log t}{\phi(n)}.$$

Whenever  $n/\phi(n)$  is bounded, as it will be whenever  $n$  is a prime, prime power or a product of two prime powers, we can replace this bound with

$$(11) \quad \sum_{\substack{p \leq t \\ p \in \mathcal{P}_n}} \frac{1}{p} \leq \frac{c \log \log t}{n}$$

for some absolute constant  $c$ . We include the  $c$  because occasionally we require an inequality as opposed to an estimate. We will also require the following asymptotic from [8, Theorem 1]

$$(12) \quad \sum_{\substack{p \in \mathcal{P}_n \\ p \leq t}} \frac{1}{p} = \frac{\log \log t}{\phi(n)} + O\left(\frac{\log n}{\phi(n)}\right),$$

which easily implies that

$$(13) \quad \sum_{\substack{p \in \mathcal{P}_n \\ p \leq t}} \frac{1}{p-1} = \frac{\log \log t}{\phi(n)} + O\left(\frac{\log n}{\phi(n)}\right),$$

since the difference is

$$\sum_{\substack{p \in \mathcal{P}_n \\ p \leq t}} \frac{1}{p(p-1)} \leq \sum_{m=1}^{\infty} \frac{1}{mn(mn+1)} < \frac{1}{n^2} \sum_{m=1}^{\infty} \frac{1}{m^2} \ll \frac{1}{n^2}.$$

### 3. REQUIRED PROPOSITIONS AND PROOF OF THEOREM 1

As mentioned previously, the main contribution to  $\log(n/\lambda_k(n))$  will come from  $\log(\phi_k(n)/\lambda_k(n))$ . Finding this term will involve a summation over prime powers which divide each of  $\phi_k(n)$  and  $\lambda_k(n)$ . It turns out that the largest contribution to this term will come from small primes which divide  $\phi_k(n)$ . By small, we mean primes  $q \leq (\log \log x)^k = y^k$ . Hence we will split the sum into small primes and large primes  $q > y^k$ . Therefore to prove Theorem 1 we will require the following propositions. The first summations deal with the large primes which divide  $\phi_k(n)$  and the second involves the large primes whose prime powers divide  $\phi_k(n)$ . We will show that the contribution of these primes to the main sum is small and hence it will end up as part of the error term.

**Proposition 3.**

$$\sum_{\substack{q > y^k \\ \nu_q(\phi_k(n))=1}} (\nu_q(\phi_k(n)) - \nu_q(\lambda_k(n))) \log q \ll y^k \psi(x)$$

for almost all  $n \leq x$ .

**Proposition 4.**

$$\sum_{\substack{q > y^k \\ \nu_q(\phi_k(n)) \geq 2}} \nu_q(\phi_k(n)) \log q \ll y^k \psi(x)$$

for almost all  $n \leq x$ .

Since the main contribution will come from small primes dividing  $\phi_k(n)$ , the next proposition will show that the contribution of small primes dividing  $\lambda_k(n)$  to the main sum can also be merged into the error term.

**Proposition 5.**

$$\sum_{q \leq y^k} \nu_q(\lambda_k(n)) \log q \ll y^k \psi(x)$$

for almost all  $n \leq x$ .

That will leave us with the contribution of small primes dividing  $\phi_k(n)$ . We will use an additive function to approximate this sum. Let  $h_k(n)$  be the additive function defined by

$$h_k(n) = \sum_{p_1|n} \sum_{p_2|p_1-1} \cdots \sum_{p_k|p_{k-1}-1} \sum_{q \leq y^k} \nu_q(p_k-1) \log q.$$

The following proposition shows that the difference between the sum involving the small primes dividing  $\phi_k(n)$  and the term  $h_k(n)$  is small.

**Proposition 6.**

$$\sum_{q \leq y^k} \nu_q(\phi_k(n)) \log q = h_k(n) + O(y^{k-1} \log y \cdot \psi(x))$$

for almost all  $n \leq x$ ,

That leaves us with  $\log(\phi_k(n)/\lambda_k(n))$  being approximated by  $h_k(n)$ . The last proposition will obtain an asymptotic formula for  $h_k(n)$ . From there we will have enough armoury to tackle Theorem 1.

**Proposition 7.**

$$h_k(n) = \frac{1}{(k-1)!} y^k \log y + O(y^k)$$

for almost all  $n \leq x$ .

*Proof of Theorem 1.* We start by breaking down the function  $\log(n/\lambda_k(n))$ .

$$\log \left( \frac{n}{\lambda_k(n)} \right) = \log \left( \frac{n}{\phi(n)} \right) + \log \left( \frac{\phi(n)}{\phi_2(n)} \right) + \cdots + \log \left( \frac{\phi_{k-1}(n)}{\phi_k(n)} \right) + \log \left( \frac{\phi_k(n)}{\lambda_k(n)} \right).$$

Using the lower bound  $\phi(m) \gg m/\log \log m$ , see [7, Theorem 2.3] we have that

$$\log \left( \frac{n}{\phi(n)} \right) + \log \left( \frac{\phi(n)}{\phi_2(n)} \right) + \cdots + \log \left( \frac{\phi_{k-1}(n)}{\phi_k(n)} \right) \ll \log \log \log n$$

and so

$$\log \left( \frac{n}{\lambda_k(n)} \right) = \log \left( \frac{\phi_k(n)}{\lambda_k(n)} \right) + O(\log \log \log n).$$

In fact we could have used a more precise estimate for  $\phi_i(n)/\phi_{i+1}(n)$  for  $i \geq 1$  which can be found in [2] but the one we used is more than good enough. Next we break down the remaining term into summations. We will break it up into small primes and large primes.

$$\begin{aligned} \log \left( \frac{\phi_k(n)}{\lambda_k(n)} \right) &= \sum_{q > y^k} (\nu_q(\phi_k(n)) - \nu_q(\lambda_k(n))) \log q + \sum_{q \leq y^k} (\nu_q(\phi_k(n)) - \nu_q(\lambda_k(n))) \log q \\ &= \sum_{\substack{q > y^k \\ \nu_q(\phi_k(n))=1}} (\nu_q(\phi_k(n)) - \nu_q(\lambda_k(n))) \log q + \sum_{\substack{q > y^k \\ \nu_q(\phi_k(n)) \geq 2}} (\nu_q(\phi_k(n)) - \nu_q(\lambda_k(n))) \log q \\ &\quad + \sum_{q \leq y^k} \nu_q(\phi_k(n)) \log q - \sum_{q \leq y^k} \nu_q(\lambda_k(n)) \log q. \end{aligned}$$

Note that if  $a \mid b$ , then  $\lambda(a) \mid \phi(b)$  since  $\lambda(a) \mid \phi(a) \mid \phi(ma)$  for any  $m$ . This quickly implies that  $\lambda_k(n)$  always divides  $\phi_k(n)$  for all  $k$  and so we get

$$0 \leq \sum_{\substack{q > y^k \\ \nu_q(\phi_k(n)) \geq 2}} (\nu_q(\phi_k(n)) - \nu_q(\lambda_k(n))) \log q \leq \sum_{\substack{q > y^k \\ \nu_q(\phi_k(n)) \geq 2}} (\nu_q(\phi_k(n))) \log q.$$

Using Propositions 3,4,5 and 6 we get

$$\log \left( \frac{n}{\lambda_k(n)} \right) = h_k(n) + O \left( y^k \psi(x) \right)$$

for almost all  $n \leq x$ . Finally by using Proposition 7 we get

$$\log \left( \frac{n}{\lambda_k(n)} \right) = \frac{1}{(k-1)!} y^k \log y + O \left( y^k \psi(x) \right)$$

for almost all  $n \leq x$ , finishing the proof of Theorem 1.  $\square$

#### 4. PRIME POWER DIVISORS OF $\phi_k(n)$

For various reasons throughout this paper, we are concerned with the number of  $n \leq x$  such that  $q^a$  can divide  $\phi_k(n)$ . We will analyze a few of those situations here:

Case 1:  $q^2 \mid n$ . Clearly the number of such  $n$  is at most  $\frac{x}{q^2}$ .

Case 2: There exists  $p_1 \in \mathcal{P}_{q^2}, p_2 \in \mathcal{P}_{p_1}, p_3 \in \mathcal{P}_{p_2}, \dots, p_l \in \mathcal{P}_{p_{l-1}}$  where  $p_l \mid n$ . By using (11) repeatedly we get that the number of such  $n$  is

$$\begin{aligned} \sum_{n \leq x} \sum_{p_1 \in \mathcal{P}_{q^2}} \sum_{p_2 \in \mathcal{P}_{p_1}} \dots \sum_{\substack{p_l \in \mathcal{P}_{p_{l-1}} \\ p_l \mid n}} 1 &= \sum_{p_1 \in \mathcal{P}_{q^2}} \sum_{p_2 \in \mathcal{P}_{p_1}} \dots \sum_{\substack{p_l \in \mathcal{P}_{p_{l-1}} \\ n \leq x \\ p_l \mid n}} 1 \\ &\ll \sum_{p_1 \in \mathcal{P}_{q^2}} \sum_{p_2 \in \mathcal{P}_{p_1}} \dots \sum_{p_l \in \mathcal{P}_{p_{l-1}}} \frac{x}{p_l} \\ &\ll \sum_{p_1 \in \mathcal{P}_{q^2}} \sum_{p_2 \in \mathcal{P}_{p_1}} \dots \sum_{p_{l-1} \in \mathcal{P}_{p_{l-2}}} \frac{xy}{p_{l-1}} \\ &\ll \sum_{p_1 \in \mathcal{P}_{q^2}} \sum_{p_2 \in \mathcal{P}_{p_1}} \frac{xy^{l-2}}{p_2} \\ &\ll \sum_{p_1 \in \mathcal{P}_{q^2}} \frac{xy^{l-1}}{p_1} \\ &\ll \frac{xy^l}{q^2} \end{aligned}$$

Now that we've taken care of any case where  $p \in \mathcal{P}_{q^2}$ , we are just left with the possibilities not containing any powers of  $q$ . Unfortunately these cases still allow for many possibilities which we will display in an array. There are lots of ways for a prime power  $q^a$  to arise in  $\phi_k(n)$  we now define various sets of primes that are involved in generating these powers of  $q$ , and we will eventually sum over all possibilities for these sets of primes. The set  $\mathcal{L}_{h,i}$  will denote a finite set of primes. To begin, the set  $\mathcal{L}_{1,2}$  will be an arbitrary finite set of primes in  $\mathcal{P}_q$  and let  $\mathcal{L}_{1,1}$  be empty. That is:

Case 3:

Level (1,2)

$$\mathcal{L}_{1,2} \subseteq \mathcal{P}_q.$$

Level (2,1) (Obtaining the primes in the previous level)

$\mathcal{L}_{2,1}$  is any set of primes with the property that for all  $p \in \mathcal{L}_{1,1} \cup \mathcal{L}_{1,2}$ , there exists a unique prime  $r \in \mathcal{L}_{2,1}$  such that  $r \in \mathcal{P}_p$ . In other words  $p$  will divide  $\phi(r)$  and hence the primes in  $\mathcal{L}_{2,1}$  will create the primes in  $\mathcal{L}_{1,1} \cup \mathcal{L}_{1,2}$ .

Level (2,2) (New primes in  $\mathcal{P}_q$ )

$$\mathcal{L}_{2,2} \subseteq \mathcal{P}_q.$$

In general for all  $1 < h \leq k$  we have for all  $p \in \mathcal{L}_{h-1,1} \cup \mathcal{L}_{h-1,2}$  there exists a unique prime  $r \in \mathcal{L}_{h,1}$  such that  $r \in \mathcal{P}_p$ ,  $\mathcal{L}_{h,2}$  is an arbitrary subset of  $\mathcal{P}_q$ , and

$$r \in \mathcal{L}_{k,1} \cup \mathcal{L}_{k,2} \Rightarrow r \mid n.$$

Some description of the terms are in order including some helpful definitions.

**Definition 8.** An incarnation  $I$  of Case 3 is some specified description of how the primes in a lower level create the primes in the level directly above.

For example, for  $k = 3$ , an incarnation  $I$  for which  $q^4 \mid \phi_3(n)$  would be  $s_1, s_2, s_3, r_3, r_4 \in \mathcal{P}_q$  where  $r_1 \in \mathcal{P}_{s_1}, r_2 \in \mathcal{P}_{s_2 s_3}, p_1 \in \mathcal{P}_{r_1 r_2}, p_2 \in \mathcal{P}_{r_3 r_4}$ , with  $p_1 p_2 \mid n$ .

**Definition 9.** A subincarnation of  $I$  is an incarnation with added conditions. In other words if  $J$  is a subincarnation of  $I$  and an integer  $n$  satisfies incarnation  $J$ , then it will also satisfy incarnation  $I$ .

For example,  $I$  is a subincarnation of the incarnation  $s_1, s_3, r_3, r_4 \in \mathcal{P}_q$  where  $r_1 \in \mathcal{P}_{s_1}, r_2 \in \mathcal{P}_{s_3}, p_1 \in \mathcal{P}_{r_1 r_2}, p_2 \in \mathcal{P}_{r_3 r_4}$ , with  $p_1 p_2 \mid n$ .

Let  $p$  be a prime in  $\mathcal{L}_{h,i}$  which we need to divide  $\phi_{k-h+1}(n)$ . The definition of  $\mathcal{L}_{h,i}$  ensures that there is a unique prime dividing  $\phi_{k-h}(n)$  for which  $p \mid r - 1$ . The primes in levels  $(k, 1), (k, 2)$  dividing  $n$  are for the base case of the recursion, so that each prime divides  $\phi_0(n) = n$ . When  $i = 2$  we are introducing new primes to get greater powers of  $q$  in  $\phi_k(n)$ . Note that it's not necessary to have any primes on the levels  $(i, 2)$ . In fact the "worst case scenario" that we will see has no primes on these except Level (1,2).

Now that we've described the way to get  $q^a \mid \phi_k(n)$ , what is our exponent  $a$ ? Let  $m_{h,i} = \#\mathcal{L}_{h,i}$ . From the recursion above we can see that  $q^{m_{k,2}} \mid \phi(n)$  and so do the primes in  $\mathcal{L}_{k-1,1}$ . For the second iteration of  $\phi$ ,  $q^{m_{k,2}-1+m_{k-1,2}} \mid \phi_2(n)$  and so do the primes in  $\mathcal{L}_{k-2,1}$ . Hence the power of  $q$  which divides  $\phi_k(n)$  is

$$(14) \quad \max_{1 \leq j \leq k} (m_{1,1} + \sum_{2 \leq h \leq j} (m_{h,2} - 1))$$

where the sum can be empty if there are no primes in the second level  $(j, 2)$  or there are not enough to survive, i.e.  $m_{j,2} < j - 1$  and hence  $q \nmid \phi_j(\prod_{\mathcal{L}_{j,2}} p)$ . Without loss of generality, we can assume the former, since the later is a subincarnation of the former.

Now we'll introduce some notation to be used in future propositions. For any single incarnation of Case 3, let  $M$  be the total number of primes,  $N$  be the total new primes introduced at the levels  $(h, 2)$  and  $H$  be the maximum necessary level  $(h, 2)$ . Specifically

$$M = \sum_h (m_{h,1} + m_{h,2}) \quad N = \sum_{h \leq H} m_{h,2}$$

and  $H$  yields the maximum value in (14). Note that under this notation,  $q^{N-H+1} \mid \phi_k(n)$ . For example, in the incarnation  $I$  above,

$$\mathcal{L}_{1,2} = \{s_1, s_2, s_3\}, \mathcal{L}_{2,1} = \{r_1, r_2\}, \mathcal{L}_{2,2} = \{r_3, r_4\}, \mathcal{L}_{3,1} = \{p_1, p_2\}, \mathcal{L}_{3,2} = \emptyset$$

as well as

$$m_{1,2} = 3, m_{2,1} = 2, m_{2,2} = 2, m_{3,1} = 2, m_{3,2} = 0.$$

Hence  $M = 9, N = 5, H = 2$  and so the power of  $q$  which divides  $\phi_3(n)$  is  $5 - 2 + 1 = 4$  as expected.

Now that we've described Case 3, how many possible  $n$  are in that case?

**Lemma 10.** *The number of  $n \leq x$  satisfying any incarnation of Case 3 is*

$$O\left(c^M \frac{xy^M}{q^N}\right)$$

where  $c$  is the constant from equation (11).

*Proof.* Let  $\mathcal{L}_h = \mathcal{L}_{h,1} \cup \mathcal{L}_{h,2}$ . We use Brun-Titchmarsh (11) for all the primes at each level of Case 3, so the number of  $n$  is

$$\begin{aligned} \sum_{n \leq x} \sum_{p_1 \in \mathcal{L}_1} \sum_{p_2 \in \mathcal{L}_2} \cdots \sum_{p_k \in \mathcal{L}_k} 1 &= \sum_{p_1 \in \mathcal{L}_1} \sum_{p_2 \in \mathcal{L}_2} \cdots \sum_{p_k \in \mathcal{L}_k} \sum_{\substack{p_k | n \\ n \leq x}} 1 \\ &\ll \sum_{p_1 \in \mathcal{L}_1} \sum_{p_2 \in \mathcal{L}_2} \cdots \sum_{p_k \in \mathcal{L}_k} \frac{x}{\prod_{p_k \in \mathcal{L}_k} p_k}. \end{aligned}$$

Note that we have repeatedly counted the same primes in the sum as we can reorder the primes in each level. It won't be important here, but will need to be more carefully addressed later. Since the primes in level  $(k, 1)$  gave us some  $p_k \in \mathcal{P}_{p_{k-1}}$  for all the primes in  $\mathcal{L}_{k-1}$ , and for  $p \in \mathcal{L}_{k,k}$  we have  $p \in \mathcal{P}_q$ . By Brun-Titchmarsh (11) we get that the above sum is

$$\ll \sum_{p_1 \in \mathcal{L}_1} \sum_{p_2 \in \mathcal{L}_2} \cdots \sum_{p_{k-1} \in \mathcal{L}_{k-1}} \frac{x(cy)^{m_{k,1}+m_{k,2}}}{\prod_{p_{k-1} \in \mathcal{L}_{k-1}} p_{k-1} q^{m_{k,2}}}.$$

Once again we get  $m_{k-1,1} + m_{k-1,2}$  new applications of Brun-Titchmarsh giving the new primes in level  $k-2$  as well as  $m_{k-1,2}$  new powers of  $q$ . Continuing along in this manner we get:

$$\begin{aligned} &\ll \sum_{p_1 \in \mathcal{L}_1} \frac{x(cy)^{\sum_{2 \leq i \leq k} (m_{i,1}+m_{i,2})}}{\prod_{p_1 \in \mathcal{L}_1} p_1 q^{\sum_{2 \leq i \leq k} m_{i,2}}} \\ &\ll \frac{x(cy)^{\sum_{1 \leq i \leq k} (m_{i,1}+m_{i,2})}}{q^{\sum_{1 \leq i \leq k} m_{i,2}}} = \frac{x(cy)^M}{q^N}. \end{aligned}$$

□

The last thing we'll consider in this section about the ways to obtain  $\phi_k(n)$  is to determine the number of possible incarnations of Case 3. We note that there are lots of incarnations which are subincarnations of others. We will develop a concept of minimality.

**Definition 11.** An incarnation of Case 3 is minimal if it does not contain any strings of  $p_1 \in \mathcal{P}_{p_2}, p_2 \in \mathcal{P}_{p_3} \dots p_{k-1} \in \mathcal{P}_{p_k}$  where  $p_k \mid n$ .

Note that any incarnation of Case 3 is a subincarnation of a minimal one. We now use this concept to show the number of necessary incarnations of Case 3 is small.



## 5. LARGE PRIMES DIVIDING $\phi_k(n)$

In this section we will prove the two propositions dealing with  $q$  being large. We'll start with the proposition where  $\nu_q(\phi_k(n)) = 1$ .

*Proof of Proposition 3.* It suffices to show

$$\sum_{n \leq x} \sum_{\substack{q > y^k \\ \nu_q(\phi_k(n))=1}} (\nu_q(\phi_k(n)) - \nu_q(\lambda_k(n))) \log q \ll xy^k$$

as then there are at most  $O\left(\frac{xy^k}{y^k \psi(x)}\right) = O\left(\frac{x}{\psi(x)}\right)$  such  $n$  where the bound for the sum in Proposition 3 fails to hold. We examine the cases where  $\nu_q(\phi_k(n)) = 1$ . Using the notation in Lemma 10 we have two subcases for Case 3, whether  $N = 1$  or  $N > 1$ .

Suppose  $N = 1$ , then  $H = 1$ ,  $m_{1,2} = 1$  and  $m_{h,2} = 0$  for  $1 < h \leq k$ . Since  $m_{h,1} \leq m_{h-1,1} + m_{h-1,2}$  we get  $m_{h,1} \leq 1$  for all  $1 \leq h \leq k$ . Hence  $m_{h,1} = 1$  for all  $h \leq k$  and so we get the case:

$$p_1 \in \mathcal{P}_q, p_2 \in \mathcal{P}_{p_1}, p_3 \in \mathcal{P}_{p_2}, \dots, p_k \in \mathcal{P}_{p_{k-1}}$$

where  $p_k \mid n$ . However in this case we also get  $\nu_q(\lambda_k(n)) = 1$  giving us no additions to our sum.

Suppose  $N > 1$ , then  $M = \sum_h (m_{h,1} + m_{h,2}) \leq k \sum_h m_{h,2} = kN$  so the number of cases we get are

$$O\left(c^M \frac{xy^M}{q^N}\right) \ll \frac{c^M xy^{kN}}{q^N} \ll \frac{c^M xy^{2k}}{q^2}$$

since  $y > q^k$ . Since  $\nu_q(\phi_k(n)) = N - H + 1$  and  $H \leq k$ ,  $N \leq k$  implying that  $M \leq k^2$ . Hence  $c^M$  is bounded as a function of  $k$ . Also since  $M$  is bounded in terms of  $k$ , there are  $O_k(1)$  possible incarnations of Case 3, and the bound already absorbs the possibilities from Cases 1 and 2. Hence we have

$$\begin{aligned} \sum_{q > y^k} \sum_{\substack{n \leq x \\ \nu_q(\phi_k(n))=1}} (\nu_q(\phi_k(n)) - \nu_q(\lambda_k(n))) \log q &\leq \sum_{q > y^k} \sum_{\substack{n \leq x \\ \nu_q(\phi_k(n))=1 \\ N > 1}} \log q \\ &\ll \sum_{q > y^k} \frac{xy^{2k} \log q}{q^2} \\ &\ll xy^k \end{aligned}$$

by (4). □

We turn our attention to  $\nu_q(\phi_k(n)) > 1$ . We have to be more careful here since we can't guarantee that the number of incarnations of Case 3 is  $O_k(1)$ . We'll start by proving a lemma which can eliminate a lot of those cases.

**Lemma 12.** *Let  $q > y^k$  and  $S_q = S_q(x)$  consist of all  $n \leq x$  such that Case 1,2 or Case 3 where  $M \leq k(N-1)$  occurs. Then*

$$\#S_q \ll \frac{xy^k}{q^2}$$

*Proof.* There are clearly  $O_k(1)$  incarnations of Cases 1 and 2 and each yield at most  $O(xy^k/q^2)$  such  $n$ . By Lemma 10 for each incarnation of Case 3, we get at most

$$O\left(\frac{c^M y^M}{q^N}\right) \ll \frac{c^M y^k}{q^2}$$

such  $n$  since  $M \leq k(N-1)$  and  $q > y^k$ . It remains to show we only require  $O_k(1)$  such incarnations. Suppose  $n$  satisfies an incarnation with  $M \leq k(N-1)$ . Then it also satisfies a minimal incarnation with  $M \leq k(N-1)$  since removing a string of  $p_1 \in \mathcal{P}_{p_2}, p_2 \in \mathcal{P}_{p_3} \dots p_{k-1} \in \mathcal{P}_{p_k}$ , would decrease  $N$  by 1 and  $M$  by  $k$  leaving the inequality unchanged. Secondly we can assume that  $n$  also satisfies an incarnation where  $k(N-2) < M \leq k(N-1)$  since we can keep eliminating primes in the  $\mathcal{L}_{i,2}$ , which decrease  $N$  by 1, but  $M$  by at most  $k$ . This must eventually produce an incarnation where  $k(N-2) < M \leq k(N-1)$  since if we eliminate all primes in the  $\mathcal{L}_{i,2}$  but 1, then  $M > k(N-1)$ . Also note that the condition  $m_{h,1} \leq m_{h-1,1} + m_{h-1,2}$  forces  $M \leq kN$ . If  $M$  is bounded between  $k(N-2)$  and  $kN$  and the incarnation is minimal, we get that  $N$  is bounded by  $2k$  since eliminating a prime in  $\mathcal{L}_{i,2}$  can only shrink  $M$  by at most  $k-1$  since our incarnation is minimal.

Therefore  $n$  satisfies an incarnation where  $N$  and hence  $M$  are bounded functions of  $k$ . Since there are only  $O_k(1)$  such incarnations, we get our result, noting that  $c^M$  can be absorbed into the constant as well.  $\square$

*Proof of Proposition 4.* Let  $S = S(x) = \bigcup_{q > y^k} S_q$ . Using Lemma 12 we have

$$\#S \leq \sum_{q > y^k} \#S_q \ll \sum_{q > y^k} \frac{xy^k}{q^2} \ll xy^k \sum_{q > y^k} \frac{1}{q^2} \ll \frac{xy^k}{\log(y^k)y^k} \ll \frac{x}{\psi(x)}$$

by (5). As for the  $n$  with  $n \notin S$  and  $a = \nu_q(\phi_k(n)) > 1$ , the only remaining case is that  $M > k(N-1)$ . Recall that  $a = N + H - 1$ . If  $H = 1$ , then  $N = m_{1,2} = a$ , and so  $m_{2,1} = a - 1$  or  $a$ . Otherwise for  $k \geq 2$ ,

$$M = \sum_h m_{h,1} \leq a + (k-1)m_{2,1} \leq a + (k-1)(a-2) = k(a-1) - k + 2 \leq (k-1)N$$

leading to a contradiction. If  $H > 1$ , then we again wish to show that  $m_{2,1} \geq a - k$ .

$$\begin{aligned} M &= \sum_h (m_{h,1} + m_{h,2}) \\ &\leq km_{1,2} + (k-1) \sum_{h>1} m_{h,2} \\ &= m_{1,2} + (k-1)N \\ &= k(N-1) - N + k + m_{1,2} \end{aligned}$$

which implies  $m_{1,2} > N - k$  and so  $\sum_{h>1} m_{h,2} = N - m_{1,1} < k$ . Therefore if  $m_{2,1} < a - k$ , then

$$\begin{aligned}
M &= \sum_h (m_{h,1} + m_{h,2}) \\
&\leq m_{1,2} + (k-1)m_{2,1} + (k-1) \sum_{h>1} m_{h,2} \leq a + (k-1)(a-k-1) + (k-1)(k-1) \\
&= ak - 2k \\
&\leq k(N-1)
\end{aligned}$$

as  $N > a$  again leading to a contradiction. Hence  $m_{2,1} \geq a - k$  and so we can get

$$\begin{aligned}
\sum_{\substack{n \notin S \\ n \leq x}} \sum_{\substack{q > y^k \\ \nu_q(\phi_k(n)) > 1}} (\nu_q(\phi_k(n)) \log q) &\leq 2 \sum_{\substack{n \notin S \\ n \leq x}} \sum_{\substack{q > y^k \\ \nu_q(\phi_k(n)) > 1}} (\nu_q(\phi_k(n)) - 1) \log q \\
&\ll \sum_{q > y^k} \log q \sum_{a \geq 2} a \sum_{\substack{n \leq x \\ n \notin S \\ \nu_q(\phi_k(n)) = a}} 1.
\end{aligned}$$

Unfortunately, just blindly using the Brun-Titchmarsh inequality in (11) won't be good enough as we must sum over all  $a$ . Let  $g(a, k) = (a - k)!$  if  $a \geq k$  or 1 otherwise and note that since we have  $m_{1,2} \geq a - k$ , we have at least  $g(a, k)$  permutations of the same primes. Then by using Lemma 10 we get

$$a \sum_{q > y^k} \log q \sum_{\substack{n \leq x \\ n \notin S \\ \nu_q(\phi_k(n)) = a}} 1 \ll a \frac{x(cy)^M}{q^N g(a, k)} \ll \frac{ac^{k(a+k-1)} xy^{2k}}{q^2 g(a, k)}$$

using the assumption that  $q > y^k$  and  $M \leq kN \leq k(a + k - 1)$ . Hence we get our sum is

$$\begin{aligned}
\sum_{\substack{n \notin S \\ n \leq x}} \sum_{\substack{q > y^k \\ \nu_q(\phi_k(n)) > 1}} (\nu_q(\phi_k(n)) \log q) &\ll \sum_{q > y^k} \log q \sum_{a \geq 2} \frac{ac^{k(a+k-1)} xy^{2k}}{q^2 g(a, k)} \\
&= xy^{2k} \sum_{q > y^k} \frac{\log q}{q^2} \sum_{a \geq 2} \frac{ac^{k(a+k-1)}}{g(a, k)}
\end{aligned}$$

However the latter sum converges to some function depending on  $k$ , and so we get

$$\ll xy^{2k} \sum_{q > y^k} \frac{\log q}{q^2} \ll xy^k$$

by (4). □

## 6. SMALL PRIMES DIVIDING $\lambda_k(n)$

We now turn our attention to the bound involving  $\lambda_k(n)$  in the summand. Just like when we were dealing with the number of cases where  $q^a \mid \phi_k(n)$ , we will need a lemma to deal with the number of cases where  $q^a \mid \lambda_k(n)$ . Fortunately this case is much simpler as the only two ways for  $q^a \mid \lambda(n)$  is for  $q^{a+1} \mid n$  or for there to exist  $p \mid n$  with  $p \in \mathcal{P}_{q^a}$ . Note that these conditions aren't sufficient, but are necessary when  $q = 2$ .

**Lemma 13.** *The number of positive integers  $n \leq x$  for which  $q^a \mid \lambda_k(n)$  is  $O(\frac{xy^k}{q^a})$ .*

*Proof.* We'll proceed by induction on  $k$ . If  $k = 1$ , then  $q^a \mid \lambda(n)$  if  $q^{a+1} \mid n$  or  $p \in \mathcal{P}_{q^a}$  with  $p \mid n$ . The number of such  $n$  is at most

$$\sum_{\substack{n \leq x \\ q^{a+1} \mid n}} 1 + \sum_{\substack{n \leq x \\ p \in \mathcal{P}_{q^a} \\ p \mid n}} 1 \ll \frac{x}{q^{a+1}} + \sum_{p \in \mathcal{P}_{q^a}} \frac{x}{p} \ll \frac{x}{q^{a+1}} + \frac{xy}{q^a} \ll \frac{xy}{q^a}.$$

using (11). Suppose the number of  $n \leq x$  for which  $q^a \mid \lambda_{k-1}(n)$  is  $O(\frac{xy^{k-1}}{q^a})$ . If  $q^a \mid \lambda_k(n)$ , then either  $q^{a+1} \mid \lambda_{k-1}(n)$  or  $p \in \mathcal{P}_{q^a}$  with  $p \mid \lambda_{k-1}(n)$ . Hence the number of such  $n$  is bounded by

$$\sum_{\substack{n \leq x \\ q^{a+1} \mid \lambda_{k-1}(n)}} 1 + \sum_{\substack{n \leq x \\ p \in \mathcal{P}_{q^a} \\ p \mid \lambda_{k-1}(n)}} 1 \ll \frac{xy^{k-1}}{q^{a+1}} + \sum_{p \in \mathcal{P}_{q^a}} \frac{xy^{k-1}}{p} \ll \frac{xy^{k-1}}{q^{a+1}} + \frac{xy^k}{q^a} \ll \frac{xy^k}{q^a}$$

as needed. □

*Proof of Proposition 5.* Like in the proof of previous propositions, we'll show

$$\sum_{n \leq x} \sum_{q \leq y^k} \nu_q(\lambda_k(n)) \log q \ll xy^k.$$

The left hand side is equal to

$$\begin{aligned} \sum_{n \leq x} \sum_{q \leq y^k} \nu_q(\lambda_k(n)) \log q &= \sum_{n \leq x} \sum_{q \leq y^k} \log q \sum_{\substack{a \in \mathbb{N} \\ q^a \mid \lambda_k(n)}} 1 \\ &\leq \sum_{n \leq x} \sum_{q \leq y^k} \log q \sum_{\substack{a \in \mathbb{N} \\ q^a \leq y^k}} 1 + \sum_{n \leq x} \sum_{q \leq y^k} \log q \sum_{\substack{a \in \mathbb{N} \\ q^a \mid \lambda_k(n) \\ q^a > y^k}} 1. \end{aligned}$$

The first sum is

$$\sum_{n \leq x} \sum_{q \leq y^k} \log q \sum_{\substack{a \in \mathbb{N} \\ q^a \leq y^k}} 1 = \sum_{n \leq x} \sum_{m \leq y^k} \Lambda(m) \ll \sum_{n \leq x} y^k \ll xy^k,$$

and by Lemma 13 and using the geometric estimate in (6) the second sum becomes

$$\sum_{n \leq x} \sum_{q \leq y^k} \log q \sum_{\substack{a \in \mathbb{N} \\ q^a \mid \lambda_k(n) \\ q^a > y^k}} 1 \ll \sum_{q \leq y^k} \log q \sum_{\substack{a \in \mathbb{N} \\ q^a > y^k}} \frac{xy^k}{q^a} \ll \sum_{q \leq y^k} \log q \frac{xy^k}{y^k} \ll xy^k.$$

□

## 7. REDUCTION TO $h_k(n)$ FOR SMALL PRIMES

The small primes dividing  $\phi_k(n)$  are what contributes to the asymptotic term of  $\log(n/\lambda_k(n))$ . In this section we show that the important case is the supersquarefree case of  $p$  dividing  $\phi_k(n)$  which is when

$$p \in \mathcal{P}_{p_1}, p_1 \in \mathcal{P}_{p_2} \dots p_{k-1} \in \mathcal{P}_{p_k}, p_k \mid n.$$

For this reason we will approximate the sum  $\sum_{q \leq y^k} v_q(\phi_k(n)) \log q$  with

$$(15) \quad h_k(n) = \sum_{p_1 \mid n} \sum_{p_2 \mid p_1 - 1} \dots \sum_{p_k \mid p_{k-1} - 1} \sum_{q \leq y^k} \nu_q(p_k - 1) \log q.$$

*Proof of Proposition 6.* For any fixed prime  $q$ , we know that

$$v_q(\phi(m)) = \max\{0, v_q(m) - 1\} + \sum_{p \mid m} v_q(p - 1),$$

which implies

$$\sum_{p \mid m} v_q(p - 1) \leq v_q(\phi(m)) \leq v_q(m) + \sum_{p \mid m} v_q(p - 1).$$

Repeated use of this inequality for  $m = \phi_l(n)$  where  $l$  ranges from  $k - 1$  to 0 yields

$$(16) \quad \begin{aligned} \sum_{p \mid \phi_{k-1}(n)} v_q(p - 1) &\leq v_q(\phi_k(n)) \\ &\leq \sum_{p \mid \phi_{k-1}(n)} v_q(p - 1) + \sum_{p \mid \phi_{k-2}(n)} v_q(p - 1) \\ &\quad + \dots + \sum_{p \mid \phi(n)} v_q(p - 1) + v_q(n). \end{aligned}$$

A prime  $p$  divides  $\phi_{k-1}(n)$  either in the supersquarefree case (ssf), or not in the supersquare-free case (nssf), yielding

$$\begin{aligned} \sum_{ssf} v_q(p - 1) &\leq \sum_{p \mid \phi_{k-1}(n)} v_q(p - 1) \\ &\leq \sum_{ssf} v_q(p - 1) + \sum_{nssf} v_q(p - 1). \end{aligned}$$

Combining this inequality with (16) yields

$$\begin{aligned} \sum_{ssf} v_q(p - 1) &\leq v_q(\phi_k(n)) \\ &\leq \sum_{ssf} v_q(p - 1) + \sum_{nssf} v_q(p - 1) + \sum_{p \mid \phi_{k-2}(n)} v_q(p - 1) + \dots + \sum_{p \mid \phi(n)} v_q(p - 1) + v_q(n). \end{aligned}$$

Subtracting the sum over the supersquarefree case, multiplying through by  $\log q$  and summing over  $q \leq y^k$  we get

$$\begin{aligned}
0 &\leq \sum_{q \leq y^k} \nu_q(\phi_k(n)) \log q - h_k(n) \\
&\leq \sum_{q \leq y^k} \sum_{nssf} v_q(p-1) \log q + \sum_{q \leq y^k} \sum_{p | \phi_{k-2}(n)} v_q(p-1) \log q + \cdots + \sum_{q \leq y^k} \sum_{p | n} v_q(p-1) \log q
\end{aligned}$$

where we get  $h_k(n)$  from (15). Hence it suffices to show that the sum on the right side becomes our error term. For the sum

$$\begin{aligned}
\sum_{n \leq x} \sum_{q \leq y^k} \sum_{p | \phi_m(n)} v_q(p-1) \log q &= \sum_{n \leq x} \sum_{q \leq y^k} \sum_{p | \phi_m(n)} \sum_{\substack{a \in \mathbb{N} \\ q^a | p-1}} \log q \\
&= \sum_{n \leq x} \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p \in \mathcal{P}_{q^a} \\ p | \phi_m(n)}} 1,
\end{aligned}$$

we'll split the sum over values of  $p \leq y^{k-1}$  and  $p > y^{k-1}$ . For  $p \leq y^{k-1}$  we uniformly get for all  $n$  that

$$\begin{aligned}
\sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p \in \mathcal{P}_{q^a} \\ p \leq y^{k-1} \\ p | \phi_m(n)}} 1 &\leq \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \pi(y^{k-1}; q^a, 1) \\
&\ll \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \frac{y^{k-1}}{\phi(q^a)} \\
&\ll y^{k-1} \sum_{q \leq y^k} \frac{\log q}{q} \\
&\ll y^{k-1} \log y
\end{aligned}$$

using the geometric estimate (6) and the prime number theorem for arithmetic progressions. As for  $p > y^{k-1}$  we fix an  $M$  and  $N$  from case 3 for which  $p \mid \phi_m(n)$ , of which there are at most  $O_k(1)$  such  $M, N$  since  $v_p(\phi(m)) = 1$ . Therefore

$$\begin{aligned}
\sum_{n \leq x} \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p > y^{k-1} \\ p \in \mathcal{P}_{q^a} \\ p | \phi_m(n)}} 1 &\ll \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p \in \mathcal{P}_{q^a} \\ p > y^{k-1}}} \frac{xy^M}{p^N} \\
&\leq \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{p \in \mathcal{P}_{q^a}} \frac{xy^{M-(k-1)(N-1)}}{p} \\
&\ll \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \frac{xy^{M-(k-1)(N-1)+1}}{q^a} \\
&\ll \sum_{q \leq y^k} \frac{xy^{M-(k-1)(N-1)+1} \log q}{q} \\
&\ll xy^{M-(k-1)(N-1)+1} \log y^k \\
&\ll xy^{M-(k-1)(N-1)+1} \log y.
\end{aligned}$$

Since the  $M, N$  were chosen for  $\phi_m(n)$  we know that  $M \leq mN$  where equality holds if and only if we are in the supersquarefree case. Now either  $m \leq k-2$  or  $m = k-1$  and we are not in the supersquarefree case. In the former case we have an error of

$$O(xy^{(k-2)N-(k-1)(N-1)+1} \log y) = O(xy^{k-N} \log y) = O(xy^{k-1} \log y)$$

since  $N \geq 1$ , or in the latter case

$$O(xy^{(k-1)N-1-(k-1)(N-1)+1} \log y) = O(xy^{k-1} \log y).$$

Thus we get

$$\begin{aligned}
\sum_{n \leq x} \left( \sum_{q \leq y^k} \sum_{n \text{ ssf}} v_q(p-1) \log q + \sum_{q \leq y^k} \sum_{p | \phi_{k-2}(n)} v_q(p-1) \log q + \dots \right. \\
\left. + \sum_{q \leq y^k} \sum_{p | n} v_q(p-1) \log q \right) \ll xy^{k-1} \log y
\end{aligned}$$

and so

$$\begin{aligned}
\sum_{q \leq y^k} \sum_{n \text{ ssf}} v_q(p-1) \log q + \sum_{q \leq y^k} \sum_{p | \phi_{k-2}(n)} v_q(p-1) \log q + \dots \\
+ \sum_{q \leq y^k} \sum_{p | n} v_q(p-1) \log q \ll y^{k-1} \log y \cdot \psi(x)
\end{aligned}$$

as required. □

## 8. REDUCTION TO THE FIRST AND SECOND MOMENTS

The Turán-Kubilius inequality [5, Lemma 3.1] asserts that if  $f(n)$  is a complex additive function, then there exists an absolute constant  $C$  such that

$$(17) \quad \sum_{n \leq x} |f(n) - M_1(x)|^2 \leq CxM_2(x)$$

where  $M_1(x) = \sum_{p \leq x} |f(p)|/p$  and  $M_2(x) = \sum_{p \leq x} |f(p)|^2/p$ . Since  $h_k(n)$  is additive we can apply this inequality where  $M_1(x) = \sum_{p \leq x} h_k(p)/p$ ,  $M_2(x) = \sum_{p \leq x} h_k(p)^2/p$ . We will need to find bounds on  $M_1$  and  $M_2$  therefore it's our goal to prove the following two propositions:

**Proposition 14.** *For all  $x > e^{e^e}$ ,*

$$M_1(x) = \frac{1}{(k-1)!} y^k \log y + O(y^k)$$

**Proposition 15.** *For all  $x > e^{e^e}$ ,*

$$M_2(x) \ll y^{2k-1} \log^{k-1} y.$$

These will lead to a proof of Proposition 7.

*Proof of Proposition 7.* Let  $N$  denote the number of  $n \leq x$  for which  $|h_k(n) - M_1(x)| > y^k$ . The contribution of such  $n$  to the sum in (17) is at least  $Ny^{2k}$ . Thus Proposition 15 implies  $N \ll x \log^{k-1} y/y$  and so Proposition 14 implies that  $h_k(n) = \frac{1}{(k-1)!} y^k \log y + O(y^k)$  except for a set of size  $O(x(\log y)^{k-1}/y)$ .  $\square$

## 9. LOTS OF SUMMATIONS

In our proofs of Propositions 14 and 15 we will see that  $M_1(x)$  and  $M_2(x)$  will reduce to summations involving  $\pi(x; p, 1)$ . We will be using some sieve techniques to bound these sums and those will require some bounds on sums on multiplicative functions involving  $\phi(m)$ . This section will involve the estimation of the latter sums.

**Lemma 16.** *For any non-negative integer  $L$  we have*

$$(18) \quad \sum_{m \leq t} \frac{m^L}{\phi(m)^{L+1}} \ll_L \log t.$$

*Proof.* If  $f(n)$  is a non-negative multiplicative function, we know that

$$(19) \quad \sum_{n \leq t} f(n) \leq \prod_{p \leq t} \sum_{r=0}^{\infty} f(p^r).$$



Applying (19) with  $\frac{m^L}{\phi(m)^{L+1}}$  yields

$$\begin{aligned}
\sum_{m \leq t} \frac{m^L}{\phi(m)^{L+1}} &\leq \prod_{p \leq t} \left( 1 + \sum_{r=1}^{\infty} \frac{p^{rL}}{(p^r - p^{r-1})^{L+1}} \right) \\
&= \prod_{p \leq t} \left( 1 + \sum_{r=1}^{\infty} \frac{p^{L-r+1}}{(p-1)^{L+1}} \right) \\
&= \prod_{p \leq t} \left( 1 + \frac{1}{(p-1)^{L+1}} \frac{p^L}{1 - \frac{1}{p}} \right) \\
&= \prod_{p \leq t} \left( 1 + \frac{p^{L+1}}{(p-1)^{L+2}} \right) \\
&\leq \exp \left( \sum_{p \leq t} \log \left( 1 + \frac{p^{L+1}}{(p-1)^{L+2}} \right) \right) \\
&= \exp \left( \sum_{p \leq t} \left( \frac{p^{L+1}}{(p-1)^{L+2}} + O_L \left( \frac{1}{p^2} \right) \right) \right) \\
&= \exp \left( \sum_{p \leq t} \left( \frac{1}{p} + O_L \left( \frac{1}{p^2} \right) \right) \right) \\
&\ll_L \log t
\end{aligned}$$

using (3). □

**Lemma 17.** *Given a positive integer  $C \leq t^\gamma$  and non-negative integer  $L$  we have*

$$(20) \quad \sum_{m \leq t} \frac{(Cm+1)^L}{\phi(Cm+1)^L \phi(m)} \ll_{L,\gamma} \log t.$$

*Proof.* It will suffice to show

$$\sum_{m \leq t} \frac{(Cm+1)^{2L-1}}{\phi(Cm+1)^{2L}} \ll_L \frac{\log t}{C}$$

as then by Cauchy–Schwarz we can get that

$$\begin{aligned}
\left( \sum_{m \leq t} \frac{(Cm+1)^L}{\phi(Cm+1)^L \phi(m)} \right)^2 &\leq \sum_{m \leq t} \frac{(Cm+1)^{2L-1}}{\phi(Cm+1)^{2L}} \sum_{m \leq t} \frac{(Cm+1)}{\phi(m)^2} \\
&\ll_L \left( \frac{\log t}{C} \right) C \log t \\
&\ll_L \log^2 t
\end{aligned}$$

by using (18). Using Mobius inversion, let  $s(n)$  be the multiplicative function defined by

$$\frac{n^{2L}}{\phi(n)^{2L}} = 1 * s = \sum_{d|n} s(d).$$

Testing at prime powers, we can easily see that

$$s(1) = 1, s(p) = \left(1 - \frac{1}{p}\right)^{-2L} - 1 \text{ and } s(p^k) = 0 \text{ for all } k \geq 2.$$

Hence

$$\begin{aligned} \sum_{m \leq t} \frac{(Cm + 1)^{2L-1}}{\phi(Cm + 1)^{2L}} &= \sum_{\substack{C < n \leq Ct+1 \\ n \equiv 1 \pmod{C}}} \frac{n^{2L-1}}{\phi(n)^{2L}} \\ &= \sum_{\substack{C < n \leq Ct+1 \\ n \equiv 1 \pmod{C}}} \frac{1}{n} \frac{n^{2L}}{\phi(n)^{2L}} \\ &= \sum_{\substack{C < n \leq Ct+1 \\ n \equiv 1 \pmod{C}}} \frac{1}{n} \sum_{d|n} s(d) \\ &= \sum_{d \leq Ct+1} s(d) \sum_{\substack{C < n \leq Ct+1 \\ d|n \\ n \equiv 1 \pmod{C}}} \frac{1}{n}. \end{aligned}$$

By (8) and noticing that  $C$  and  $d$  are relatively prime we get

$$\sum_{\substack{C < n \leq Ct+1 \\ d|n \\ n \equiv 1 \pmod{C}}} \frac{1}{n} \ll \frac{1}{C+1} + \frac{\log t}{dC}$$

where the first term occurs only if  $d \mid C+1$ . We require some estimates on  $s(d)$ .

$$\begin{aligned} \sum_{d \leq Ct+1} \frac{s(d)}{d} &\leq \prod_{p \leq Ct+1} \left(1 + \frac{(1 - 1/p)^{-2L} - 1}{p}\right) \\ &\leq \prod_{p \leq Ct+1} \left(1 + \frac{C_L}{p^2}\right) \\ &= \exp\left(\sum_{p \leq Ct+1} \log\left(1 + \frac{C_L}{p^2}\right)\right) \\ &= \exp\left(\sum_{p \leq Ct+1} O_L\left(\frac{1}{p^2}\right)\right) \\ &= \exp(O_L(1)) \\ &\ll_L 1 \end{aligned}$$

and

$$\begin{aligned}
\sum_{\substack{d \leq Ct+1 \\ d|C+1}} s(d) &\leq \sum_{d|C+1} s(d) \\
&= (1 * s)(C+1) \\
&= \left( \frac{C+1}{\phi(C+1)} \right)^{2L} \\
&\ll (\log \log C)^{2L} \\
&\ll_{\gamma} (\log \log t)^{2L} \\
&\ll_{L,\gamma} \log t.
\end{aligned}$$

Therefore

$$\sum_{m \leq t} \frac{(Cm+1)^{2L-1}}{\phi(Cm+1)^{2L}} \ll \sum_{\substack{d \leq Ct+1 \\ d|C+1}} \frac{s(d)}{C+1} + \sum_{d \leq t} \frac{s(d) \log t}{Cd} \ll_{L,\gamma} \frac{\log t}{C}$$

as needed.  $\square$

**Lemma 18.** *For positive integers  $C_1, C_2, \dots, C_r \leq t^\gamma$  and non-negative integers  $L_1, L_2, \dots, L_r$  we have*

$$(21) \quad \sum_{m \leq t} \frac{(C_1m+1)^{L_1} (C_2m+1)^{L_2} \dots (C_rm+1)^{L_r}}{\phi(C_1m+1)^{L_1} \phi(C_2m+1)^{L_2} \dots \phi(C_rm+1)^{L_r} \phi(m)} \ll_{L_1, \dots, L_r, \gamma} \log t.$$

*Proof.* We proceed by induction. The case  $r = 1$  is covered by Lemma 17. Suppose

$$\sum_{m \leq t} \frac{(C_1m+1)^{L_1} (C_2m+1)^{L_2} \dots (C_rm+1)^{L_r}}{\phi(C_1m+1)^{L_1} \phi(C_2m+1)^{L_2} \dots \phi(C_rm+1)^{L_r} \phi(m)} \ll_{L_1, \dots, L_r, \gamma} \log t.$$

By Cauchy–Schwarz, we get that

$$\begin{aligned}
&\left( \sum_{m \leq t} \frac{(C_1m+1)^{L_1} (C_2m+1)^{L_2} \dots (C_{r+1}m+1)^{L_{r+1}}}{\phi(C_1m+1)^{L_1} \phi(C_2m+1)^{L_2} \dots \phi(C_{r+1}m+1)^{L_{r+1}} \phi(m)} \right)^2 \\
&\leq \sum_{m \leq t} \frac{(C_1m+1)^{2L_1} (C_2m+1)^{2L_2} \dots (C_rm+1)^{2L_r}}{\phi(C_1m+1)^{2L_1} \phi(C_2m+1)^{2L_2} \dots \phi(C_rm+1)^{2L_r} \phi(m)} \sum_{m \leq t} \frac{(C_{r+1}m+1)^{2L_{r+1}}}{\phi(C_{r+1}m+1)^{2L_{r+1}} \phi(m)} \\
&\ll_{L_1, \dots, L_{r+1}, \gamma} \log^2 t
\end{aligned}$$

by Lemma 17, completing the proof.  $\square$

**Lemma 19.** *For positive integers  $C_1, C_2, \dots, C_r \leq t^\gamma$  and non-negative integers  $L_1, L_2, \dots, L_r, L$  we have*

$$(22) \quad \sum_{m \leq t} \frac{(C_1m+1)^{L_1} (C_2m+1)^{L_2} \dots (C_rm+1)^{L_r} m^{L-1}}{\phi(C_1m+1)^{L_1} \phi(C_2m+1)^{L_2} \dots \phi(C_rm+1)^{L_r} \phi(m)^L} \ll_{L_1, \dots, L_r, L, \gamma} \log t.$$

*Proof.* Once again we'll use Cauchy–Schwarz and the previous lemmas.

$$\begin{aligned}
& \left( \sum_{m \leq t} \frac{(C_1 m + 1)^{L_1} (C_2 m + 1)^{L_2} \dots (C_r m + 1)^{L_r} m^{L-1}}{\phi(C_1 m + 1)^{L_1} \phi(C_2 m + 1)^{L_2} \dots \phi(C_r m + 1)^{L_r} \phi(m)^L} \right)^2 \\
& \leq \sum_{m \leq t} \frac{(C_1 m + 1)^{2L_1} (C_2 m + 1)^{2L_2} \dots (C_r m + 1)^{2L_r}}{\phi(C_1 m + 1)^{2L_1} \phi(C_2 m + 1)^{2L_2} \dots \phi(C_r m + 1)^{2L_r} \phi(m)} \sum_{m \leq t} \frac{m^{2L-2}}{\phi(m)^{2L-1}} \\
& \ll_{L_1, \dots, L_r, L, \gamma} \log^2 t
\end{aligned}$$

by Lemmas 16 and 18. □

## 10. MORE SUMMATIONS INVOLVING $\pi(t, p, 1)$

The previous section involved lemmas required to prove summations including terms such as  $\pi(t, p, 1)$ . A lot of these summations will involve sieving techniques. This section will be split into proofs of two lemmas involving the summations required for the sums arising from the Propositions 14 and 15.

**Lemma 20.** *Let  $b, k, l$  be positive integers with  $2 \leq l \leq k$ . Let  $t > e^e$  be a real number and let constants  $\alpha, \alpha_1, \alpha_2$  satisfy  $0 < \alpha < 1/2$  and  $0 < \alpha_1 < \alpha_2 < 1/2$ .*

(a) *If  $b > t^\alpha$ , then*

$$(23) \quad \sum_{p_k \in \mathcal{P}_b} \sum_{p_{k-1} \in \mathcal{P}_{p_k}} \dots \sum_{p_2 \in \mathcal{P}_{p_3}} \pi(t; p_2, 1) \ll \frac{t \log t (\log \log t)^{k-2}}{b}.$$

(b) *If  $b \leq t^{\alpha_1}$ , then*

$$(24) \quad \sum_{\substack{p_l \in \mathcal{P}_b \\ p_l > t^{\alpha_2}}} \sum_{p_{l-1} \in \mathcal{P}_{p_l}} \dots \sum_{p_2 \in \mathcal{P}_{p_3}} \pi(t; p_2, 1) \ll \frac{b^{l-1} t}{\phi(b)^l \log t}.$$

(c) *If  $b \leq t^{\alpha_1}$ , then*

$$(25) \quad \sum_{p_l \in \mathcal{P}_b} \sum_{p_{l-1} \in \mathcal{P}_{p_l}} \dots \sum_{p_2 \in \mathcal{P}_{p_3}} \pi(t; p_2, 1) \ll \frac{t (\log \log t)^{l-1}}{\phi(b) \log t}.$$

*The implicit constants in (a) – (c) depend on the choices of the  $\alpha$ .*

*Proof.* For (23) we just use the trivial estimate  $\pi(t; p_2, 1) \leq t/p_2$  and several uses of Brun-Titchmarsh (11) to get

$$\begin{aligned}
\sum_{p_k \in \mathcal{P}_b} \sum_{p_{k-1} \in \mathcal{P}_k} \cdots \sum_{p_2 \in \mathcal{P}_3} \pi(t; p_2, 1) &\leq \sum_{p_k \in \mathcal{P}_b} \sum_{p_{k-1} \in \mathcal{P}_k} \cdots \sum_{p_2 \in \mathcal{P}_3} \frac{t}{p_2} \\
&\ll t \sum_{p_k \in \mathcal{P}_b} \sum_{p_{k-1} \in \mathcal{P}_k} \cdots \sum_{p_3 \in \mathcal{P}_4} \frac{\log \log t}{p_3} \\
&\ll t \sum_{p_k \in \mathcal{P}_b} \frac{(\log \log t)^{k-2}}{p_k} \\
&\leq t \sum_{\substack{m \equiv 1 \pmod{b} \\ t^\alpha \leq m \leq t}} \frac{(\log \log t)^{k-2}}{m} \\
&\leq \frac{t \log t (\log \log t)^{k-2}}{b}
\end{aligned}$$

where  $m > 1$  and  $m \equiv 1 \pmod{b}$  imply that  $m > b$  and by using (7). As for (24) we get

$$\begin{aligned}
&\sum_{\substack{p_l \in \mathcal{P}_b \\ l > t^{\alpha_2}}} \sum_{p_{l-1} \in \mathcal{P}_l} \cdots \sum_{p_2 \in \mathcal{P}_3} \pi(t; p_2, 1) \\
&= \sum_{\substack{p_l \in \mathcal{P}_b \\ l > t^{\alpha_2}}} \sum_{p_{l-1} \in \mathcal{P}_l} \cdots \sum_{p_3 \in \mathcal{P}_4} \#\{(m_1, p_2) : p_2 = 1 \pmod{p_3}, p_2 > t^{\alpha_2}, m_1 p_2 + 1 \leq t, p_2, m_1 p_2 + 1 \text{ prime}\} \\
&= \sum_{\substack{p_l \in \mathcal{P}_b \\ l > t^{\alpha_2}}} \sum_{p_{l-1} \in \mathcal{P}_l} \cdots \sum_{p_4 \in \mathcal{P}_5} \#\{(m_1, m_2, p_3) : p_3 = 1 \pmod{p_4}, p_3 > t^{\alpha_2}, m_1(m_2 p_3 + 1) + 1 \leq t, \\
&\quad \{p_3, m_2 p_3 + 1, m_1(m_2 p_3 + 1) + 1\} \text{ prime}\} \\
&= \#\{(m_1, m_2, \dots, m_{l-1}, p_l) : p_l = 1 \pmod{b}, p_l > t^{\alpha_2}, m_1(m_2 \dots (m_{l-2}(m_{l-1} p_l + 1) + 1) + \dots \\
&\quad + 1 \leq t, \{p_l, m_{l-1} p_l + 1, m_{l-2}(m_{l-1} p_l + 1) + 1, \dots, m_1(m_2 \dots (m_{l-2}(m_{l-1} p_l + 1) + 1) \\
&\quad + \dots + 1\} \text{ prime}\} \\
&\leq \sum_{m_1 \dots m_{l-1} \leq t^{1-\alpha_2}} \#\{p_l < t/m_1 \dots m_{l-1} : p_l = 1 \pmod{b}, \\
&\quad \{p_l, m_{l-1} p_l + 1, m_{l-2}(m_{l-1} p_l + 1) + 1, \dots, m_1(m_2 \dots (m_{l-2}(m_{l-1} p_l + 1) + 1) + \dots + 1\} \text{ prime}\}.
\end{aligned}$$

From here will need to use Brun's Sieve method (see [4, Theorem 2.4]) to get that

$$\begin{aligned}
&\#\{p_l < t/m_1 \dots m_{l-1} : p_l = 1 \pmod{b}, \\
&\quad \{p_l, m_{l-1} p_l + 1, m_{l-2}(m_{l-1} p_l + 1) + 1, \dots, m_1(m_2 \dots (m_{l-2}(m_{l-1} p_l + 1) + 1) + \dots + 1\} \text{ prime}\} \\
&\ll \frac{E^{l-1}}{\phi(E)^{l-1}} \frac{b^{l-1}}{\phi(b)^{l-1}} \frac{bc_1 \dots c_{l-1}}{\phi(bc_1 \dots c_{l-1})} \frac{t/m_1 \dots m_{l-1} b}{(\log t/m_1 \dots m_{l-1} b)^l}
\end{aligned}$$

where the  $c_i$  and  $E$  are

$$E = \left( \prod_{i=1}^{l-1} m_i^{i(i+1)/2} \right) (1 + m_1 + m_1 m_2 + \cdots + m_1 \dots m_{l-3}) (1 + m_2 + m_2 m_3 + \cdots + m_2 \dots m_{l-4}) \\ \dots (1 + m_{l-3}) (1 + m_1 + m_1 m_2 + \cdots + m_1 \dots m_{l-4}) (1 + m_2 + m_2 m_3 + \cdots + m_2 \dots m_{l-4}) \\ \dots (1 + m_{l-4}) \dots (1 + m_1)$$

and for  $1 \leq i \leq l-1$ ,

$$c_i = 1 + m_i + m_i m_{i+1} + \cdots + m_i \dots m_{l-2}, c_{l-1} = 1.$$

Now using  $\phi(mn) \leq \phi(m)\phi(n)$  and  $m_1 \dots m_{l-1} b \leq t^{1+\alpha_1-\alpha_2}$  where  $1 + \alpha_1 - \alpha_2 < 1$  we get

$$\ll \frac{E^{l-1}}{\phi(E)^{l-1}} \frac{b^{l-1}}{\phi(b)^l} \frac{c_1}{\phi(c_1)} \cdots \frac{c_{l-1}}{\phi(c_{l-1})} \frac{t}{m_1 \dots m_{l-1} (\log t)^l}.$$

Using

$$\frac{m^L}{\phi(m^L)} = \frac{m}{\phi(m)},$$

we get the sum

$$\sum_{m_1 \dots m_{l-1} \leq t^{1-\alpha_2}} \frac{E^{l-1}}{\phi(E)^{l-1}} \frac{c_1}{\phi(c_1)} \cdots \frac{c_{l-1}}{\phi(c_{l-1})} \frac{1}{m_1 \dots m_{l-1}} = \sum_{m_1 \dots m_{l-1} \leq t^{1-\alpha_2}} \frac{(E^*)^{l-1}}{\phi(E^*)^{l-1}} \frac{c_1}{\phi(c_1)} \cdots \frac{c_{l-1}}{\phi(c_{l-1})} \frac{1}{m_1 \dots m_{l-1}}$$

where

$$E^* = (1 + m_1 + m_1 m_2 + \cdots + m_1 \dots m_{l-3}) (1 + m_2 + m_2 m_3 + \cdots + m_2 \dots m_{l-4}) \\ \dots (1 + m_{l-3}) (1 + m_1 + m_1 m_2 + \cdots + m_1 \dots m_{l-4}) (1 + m_2 + m_2 m_3 + \cdots + m_2 \dots m_{l-4}) \\ \dots (1 + m_{l-4}) \dots (1 + m_1).$$

We have that every factor in  $E^*$  as well as the  $c_i$  are of the form  $1 + Cm_i$  for some  $i$  or of the form  $m_i^L$ . Hence using  $l-1$  applications of Lemmas 16, 18 or 19 we can pick off the factors of the form  $(1 + Cm_i)$  one at a time.

$$\sum_{m_1 \dots m_{l-1} \leq t^{1-\alpha_2}} \frac{E^{l-1}}{\phi(E)^{l-1}} \frac{c_1}{\phi(c_1)} \cdots \frac{c_{l-1}}{\phi(c_{l-1})} \frac{1}{m_1 \dots m_{l-1}} \\ \ll \sum_{m_2 \dots m_{l-1} \leq t^{1-\alpha_2}} \frac{(E')^{l-1}}{\phi(E')^{l-1}} \frac{c'_1}{\phi(c'_1)} \cdots \frac{c'_{l-1}}{\phi(c'_{l-1})} \frac{1}{m_2 \dots m_{l-1}} (\log t) \\ \ll \sum_{m_3 \dots m_{l-1} \leq t^{1-\alpha_2}} \frac{(E'')^{l-1}}{\phi(E'')^{l-1}} \frac{c''_1}{\phi(c''_1)} \cdots \frac{c''_{l-1}}{\phi(c''_{l-1})} \frac{1}{m_3 \dots m_{l-1}} (\log^2 t) \\ \ll \cdots \ll (\log t)^{l-1}.$$

where the  $E^{(e)}, c_i^{(e)}$  denote the  $E^*$  and  $c_i$  terms with the factors of the form  $1 + Cm_1$  through  $1 + Cm_e$  removed. Note that the  $C$  are at most  $1 + t + t^2 + \cdots + t^{k-3} \leq t^{k-2}$  and  $l \leq k$  so

the implied constant only depends on  $k$ . Therefore

$$\sum_{\substack{p_l \in \mathcal{P}_b \\ l > t^{\alpha_2}}} \sum_{p_{l-1} \in \mathcal{P}_l} \cdots \sum_{p_2 \in \mathcal{P}_3} \pi(t; p_2, 1) \ll \frac{tb^{l-1}}{\phi(b)^l (\log t)^l} (\log t)^{l-1} = \frac{tb^{l-1}}{\phi(b)^l \log t}.$$

As for part (c), first note that  $b/\phi(b) \ll \log \log b$ , so for  $p_l > t^{\alpha_2}$ , we get that part (b) implies our bound. As for  $p_l \leq t^{\alpha_2}$  we'll split it into cases where  $p_3$  is less than or greater than  $t^{\alpha_2}$ . If  $p_3 \leq t^{\alpha_2}$ , then

$$\begin{aligned} \sum_{\substack{p_l \in \mathcal{P}_b \\ p_l \leq t^{\alpha_2}}} \sum_{p_{l-1} \in \mathcal{P}_l} \cdots \sum_{\substack{p_2 \in \mathcal{P}_3 \\ p_2 \leq t^{\alpha_2}}} \pi(t; p_2, 1) &\ll \sum_{\substack{p_l \in \mathcal{P}_b \\ p_l \leq t^{\alpha_2}}} \sum_{p_{l-1} \in \mathcal{P}_l} \cdots \sum_{\substack{p_2 \in \mathcal{P}_3 \\ p_2 \leq t^{\alpha_2}}} \frac{t}{\phi(p_2) \log t / p_2} \\ &\ll \sum_{\substack{p_l \in \mathcal{P}_b \\ p_l \leq t^{\alpha_2}}} \sum_{p_{l-1} \in \mathcal{P}_l} \cdots \sum_{\substack{p_2 \in \mathcal{P}_3 \\ p_2 \leq t^{\alpha_2}}} \frac{t}{p_2 \log t} \\ &\ll \sum_{p_l \in \mathcal{P}_b} \frac{t(\log \log t)^{l-2}}{p_l \log t} \\ &\ll \frac{t(\log \log t)^{l-1}}{\phi(b) \log t} \end{aligned}$$

If  $p_3 > t^{\alpha_2}$ , then since  $b \leq t^{\alpha_2}$  there is a minimum  $m$  such that  $p_m \leq t^{\alpha_2}$ . So using part (b) with  $l = m$  we get

$$\begin{aligned} \sum_{\substack{p_l \in \mathcal{P}_b \\ p_l \leq t^{\alpha_2}}} \sum_{p_{l-1} \in \mathcal{P}_l} \cdots \sum_{\substack{p_2 \in \mathcal{P}_3 \\ p_2 > t^{\alpha_2}}} \pi(t; p_2, 1) &\ll \sum_{\substack{p_l \in \mathcal{P}_b \\ p_l \leq t^{\alpha_2}}} \sum_{p_{l-1} \in \mathcal{P}_l} \cdots \sum_{p_{m+1} \in \mathcal{P}_{m+2}} \frac{(p_{m-1})^{m-1} t}{\phi(p_{m-1})^m \log t} \\ &\ll \sum_{\substack{p_l \in \mathcal{P}_b \\ p_l \leq t^{\alpha_2}}} \sum_{p_{l-1} \in \mathcal{P}_l} \cdots \sum_{p_{m+1} \in \mathcal{P}_{m+2}} \frac{t}{p_{m-1} \log t} \\ &\ll \frac{t(\log \log t)^{l-m}}{\phi(b) \log t} \\ &\ll \frac{t(\log \log t)^{l-1}}{\phi(b) \log t} \end{aligned}$$

since  $m \geq 2$  and by using Brun-Titchmarsh (11) which finishes part (c) and the lemma.  $\square$

As for the summations requires for the second moment, we'll note that we need twice as many sums due to  $h_k(p)^2$ . However the techniques required are similar.

**Lemma 21.** *Let  $t > e^e$  and  $0 < 2\alpha_1 < \alpha_2 < 1/2$ . Then*

(a) *If  $b_1 > t^{\alpha_1}$  or  $b_2 > t^{\alpha_1}$  then*

$$(26) \quad \sum_{\substack{p_2 \in \mathcal{P}_{b_1} \\ r_2 \in \mathcal{P}_{b_2}}} \pi(t; p_2 r_2, 1) \ll \frac{t \log^2 t}{b_1 b_2}.$$

(b) If neither  $b_1$  nor  $b_2$  exceeds  $t^{\alpha_1}$ , then

$$(27) \quad \sum_{\substack{p_k \in \mathcal{P}_{b_1} \\ r_k \in \mathcal{P}_{b_2} \\ p_k r_k > t^{\alpha_2}}} \dots \sum_{\substack{p_2 \in \mathcal{P}_{p_3} \\ r_2 \in \mathcal{P}_{r_3}}} \pi(t; p_2 r_2, 1) \ll \frac{t(\log \log t)^{k-1} b_2^{k-1}}{\phi(b_1) \phi(b_2)^k \log t} + \frac{t(\log \log t)^{k-1} b_1^{k-1}}{\phi(b_2) \phi(b_1)^k \log t}.$$

(c) If neither  $b_1$  nor  $b_2$  exceeds  $t^{\alpha_1}$ , then

$$(28) \quad \sum_{\substack{p_k \in \mathcal{P}_{b_1} \\ r_k \in \mathcal{P}_{b_2}}} \dots \sum_{\substack{p_2 \in \mathcal{P}_{p_3} \\ r_2 \in \mathcal{P}_{r_3}}} \pi(t; p_2 r_2, 1) \ll \frac{t(\log \log t)^{2k-2}}{\phi(b_1) \phi(b_2) \log t}.$$

(d) If neither  $b_1$  nor  $b_2$  exceeds  $t^{\alpha_1}$ , then

$$(29) \quad \sum_{\substack{p_k \in \mathcal{P}_{b_1} \\ r_k \in \mathcal{P}_{b_2}}} \dots \sum_{\substack{p_3 \in \mathcal{P}_{p_4} \\ r_3 \in \mathcal{P}_{r_4}}} \sum_{s \in \mathcal{P}_{p_3} \cap \mathcal{P}_{r_3}} \pi(t; s, 1) \ll \frac{t(\log \log t)^{2k-2}}{\phi(b_1) \phi(b_2) \log t}.$$

Again the implicit constants depend on our choice of the  $\alpha$ .

*Proof.* (a) is similar to part (a) of Lemma 20. For part (b) we first assume that  $p_k \leq r_k$ , then

$$\begin{aligned} & \sum_{\substack{p_k \in \mathcal{P}_{b_1} \\ r_k \in \mathcal{P}_{b_2} \\ p_k \leq r_k \\ p_k r_k > t^{\alpha_2}}} \dots \sum_{\substack{p_2 \in \mathcal{P}_{p_3} \\ r_2 \in \mathcal{P}_{r_3}}} \pi(t; p_2 r_2, 1) \\ &= \sum_{\substack{p_k \in \mathcal{P}_{b_1} \\ r_k \in \mathcal{P}_{b_2} \\ p_k r_k > t^{\alpha_2}}} \dots \sum_{\substack{p_3 \in \mathcal{P}_{p_4} \\ r_3 \in \mathcal{P}_{r_4}}} \#\{(m_1, p_2, r_2) : p_2 = 1 \pmod{p_3}, r_2 = 1 \pmod{r_3}, r_2 p_2 > t^{\alpha_2}, \\ & \quad m_1 r_2 p_2 + 1 \leq t, p_2, m_1 r_2 p_2 + 1 \text{ prime}\} \\ &= \sum_{\substack{p_k \in \mathcal{P}_{b_1} \\ p_k \leq r_k}} \sum_{p_{k-1} \in \mathcal{P}_k} \dots \sum_{\substack{p_2 \in \mathcal{P}_3 \\ p_k r_k > t^{\alpha_2}}} \sum_{r_k \in \mathcal{P}_{b_2}} \sum_{r_{k-1} \in \mathcal{P}_{r_k}} \dots \sum_{r_4 \in \mathcal{P}_{r_5}} \#\{(m_1, m_2, r_3) : r_3 = 1 \pmod{r_4}, \\ & \quad r_3 p_2 > t^{\alpha_2}, m_1 p_2 (m_2 r_3 + 1) + 1 \leq t, \{r_3, m_2 r_3 + 1, m_1 p_2 (m_2 r_3 + 1) + 1\} \text{ prime}\} \\ &= \sum_{\substack{p_k \in \mathcal{P}_{b_1} \\ p_k \leq r_k}} \sum_{p_{k-1} \in \mathcal{P}_k} \dots \sum_{p_2 \in \mathcal{P}_3} \#\{(m_1, m_2, \dots, m_{l-1}, r_l) : r_l = 1 \pmod{b_2}, p_2 r_k > t^{\alpha_2}, \\ & \quad m_1 p_2 (m_2 \dots (m_{k-2} (m_{k-1} r_k + 1) + 1) + \dots + 1 \leq t, \{r_k, m_{k-1} r_k + 1, \\ & \quad m_{k-2} (m_{k-1} r_k + 1) + 1, \dots, \\ & \quad m_1 p_2 (m_2 \dots (m_{k-2} (m_{k-1} r_k + 1) + 1) + \dots + 1\} \text{ prime}\} \\ &\leq \sum_{m_1 \dots m_{l-1} \leq t^{1-\alpha_2}} \sum_{\substack{p_k \in \mathcal{P}_{b_1} \\ p_k \leq r_k}} \sum_{p_{k-1} \in \mathcal{P}_k} \dots \sum_{p_2 \in \mathcal{P}_3} \#\{r_k < t/p_2 m_1 \dots m_{k-1} : r_k = 1 \pmod{b_2}, \\ & \quad \{r_k, m_{k-1} r_k + 1, m_{k-2} (m_{k-1} r_k + 1) + 1, \dots, \\ & \quad p_2 m_1 (m_2 \dots (m_{k-2} (m_{k-1} r_k + 1) + 1) + \dots + 1\} \text{ prime}\} \end{aligned}$$



Just like in Lemma 20 we use Brun's Sieve. However, notice that we have almost the same set, except with  $m_1$  replaced with  $m_1 p_2$ . Hence we have

$$\begin{aligned} & \#\{r_k < t/p_2 m_1 \cdots m_{k-1} : r_k \equiv 1 \pmod{b_1}, \{r_k, m_{k-1}r_k + 1, m_{k-2}(m_{k-1}r_k + 1) + 1, \\ & \quad \dots, p_2 m_1(m_2 \dots (m_{k-2}(m_{k-1}r_k + 1) + 1) + \dots + 1\} \text{ prime}\} \\ & \ll \frac{E^{k-1}}{\phi(E)^{k-1}} \frac{b_2^{k-1}}{\phi(b_2)^{k-1}} \frac{b_2 c_1 \dots c_{k-1}}{\phi(b_2 c_1 \dots c_{k-1})} \frac{t/p_2 m_1 \dots m_{k-1} b_2}{(\log t/p_2 m_1 \dots m_{k-1} b_2)^k} \end{aligned}$$

where the  $c_i$  and  $E$  are

$$\begin{aligned} E = & p_2 \left( \prod_{i=1}^{l-1} m_i^{i(i+1)/2} \right) (1 + p_2 m_1 + p_2 m_1 m_2 + \dots + p_2 m_1 \dots m_{k-3}) (1 + m_2 + m_2 m_3 + \dots \\ & + m_2 \dots m_{k-3}) \dots (1 + m_{k-3}) (1 + p_2 m_1 + p_2 m_1 m_2 + \dots + p_2 m_1 \dots m_{k-4}) \\ & (1 + m_2 + m_2 m_3 + \dots + m_2 \dots m_{k-4}) \dots (1 + m_{k-4}) \dots (1 + p_2 m_1) \end{aligned}$$

and for  $2 \leq i \leq k-1$ ,

$$\begin{aligned} c_1 &= 1 + p_2 m_1 + p_2 m_1 m_2 + \dots + p_2 m_1 \dots m_{k-2}, \\ c_i &= 1 + m_i + m_i m_{i+1} + \dots + m_i \dots m_{k-2}, c_{k-1} = 1. \end{aligned}$$

By the same methods as Lemma 20 and using that  $p_2/\phi(p_2)$  is bounded and noting that

$$\frac{t}{p_2 m_1 \dots m_{k-1} b_2} > \frac{r_k}{b_1} > t^{\alpha_2/2 - \alpha_1} = t^\epsilon$$

for some  $\epsilon > 0$  since  $\alpha_2 > 2\alpha_1$ , we get that

$$\begin{aligned} \sum_{\substack{p_k \in \mathcal{P}_{b_1} \\ r_k \in \mathcal{P}_{b_2} \\ p_k r_k > t^{\alpha_2}}} \dots \sum_{\substack{p_2 \in \mathcal{P}_{p_3} \\ r_2 \in \mathcal{P}_{r_3}}} \pi(t; p_2 r_2, 1) & \ll \frac{t b_2^{k-1}}{\phi(b_2)^k \log t} \sum_{p_k \in \mathcal{P}_{b_1}} \sum_{p_{k-1} \in \mathcal{P}_k} \dots \sum_{p_2 \in \mathcal{P}_3} \frac{1}{p_2} \\ & \ll \frac{t b_2^{k-1}}{\phi(b_2)^k \log t} \sum_{p_k \in \mathcal{P}_{b_1}} \frac{(\log \log t)^{k-2}}{p_k} \\ & \ll \frac{t (\log \log t)^{k-1} b_2^{k-1}}{\phi(b_1) \phi(b_2)^k \log t}. \end{aligned}$$

The case for  $r_k \leq p_k$  is similar. As for part (c), first note that  $b_i/\phi(b_i) \ll \log \log b_i$  for  $i \in \{1, 2\}$ . taking care of the case where  $p_k r_k > t^{\alpha_2}$ . As for  $p_k r_k \leq t^{\alpha_2}$  we get

$$\begin{aligned}
\sum_{\substack{p_k \in \mathcal{P}_{b_1} \\ r_k \in \mathcal{P}_{b_2} \\ p_k r_k \leq t^{\alpha_2}}} \cdots \sum_{\substack{p_2 \in \mathcal{P}_{p_3} \\ r_2 \in \mathcal{P}_{r_3}}} \pi(t; p_2 r_2, 1) &\ll \sum_{\substack{p_k \in \mathcal{P}_{b_1} \\ r_k \in \mathcal{P}_{b_2} \\ p_k r_k \leq t^{\alpha_2}}} \cdots \sum_{\substack{p_2 \in \mathcal{P}_{p_3} \\ r_2 \in \mathcal{P}_{r_3}}} \frac{t}{\phi(p_2 r_2) \log t / p_2 r_2} \\
&\ll \sum_{\substack{p_k \in \mathcal{P}_{b_1} \\ r_k \in \mathcal{P}_{b_2} \\ p_k r_k \leq t^{\alpha_2}}} \cdots \sum_{\substack{p_2 \in \mathcal{P}_{p_3} \\ r_2 \in \mathcal{P}_{r_3}}} \frac{t}{p_2 r_2 \log t} \\
&\ll \sum_{\substack{p_k \in \mathcal{P}_{b_1} \\ r_k \in \mathcal{P}_{b_2} \\ p_k r_k \leq t^{\alpha_2}}} \frac{t(\log \log t)^{2k-4}}{p_k r_k \log t} \\
&\ll \frac{t(\log \log t)^{2k-2}}{\phi(b_1)\phi(b_2) \log t}
\end{aligned}$$

using Brun-Titchmarsh, (11) finishing part (c). As for part (d) we note that

$$\begin{aligned}
&\sum_{\substack{p_3 \in \mathcal{P}_{p_4} \\ r_3 \in \mathcal{P}_{r_4}}} \sum_{s \in \mathcal{P}_{p_3} \cap \mathcal{P}_{r_3}} \pi(t; s, 1) \\
&= \sum_{\substack{p_3 \in \mathcal{P}_{p_4} \\ r_3 \in \mathcal{P}_{r_4}}} \#\{(m_1, s) : s = 1 \pmod{p_3 r_3}, m_1 s + 1 \leq t, s, m_1 s + 1 \text{ prime}\} \\
&= \sum_{p_3 \in \mathcal{P}_{p_4}} \#\{(m_1, m_2, r_3) : r_3 = 1 \pmod{r_4}, m_1(m_2 p_3 r_3 + 1) + 1 \leq t, \\
&\quad \{m_2 p_3 r_3 + 1, m_1(m_2 p_3 r_3 + 1) + 1 \text{ prime}\}
\end{aligned}$$

and so on, yielding a similar sieve as part (b). □

## 11. REDUCTION OF $\sum h_k(p)$ TO SMALL VALUES OF $p_k$

We will be using Euler Summation on the sum  $\sum_{p \leq t} h_k(p)$  in our efforts to find our estimate for  $M_1(x)$ . It will turn out that the large primes do not contribute much to the sum. The sum will involve estimating  $\pi(t; p, 1)$  by  $\text{li}(t)/p - 1$ . The following lemma will deal with those errors and will involve the Bombieri-Vinogradov Theorem.

**Lemma 22.** For all  $2 \leq l \leq k$ ,  $x > e^{e^e}$  and  $v > e^e$ ,

$$\begin{aligned} & \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} \left( \pi(v, p_{k-l+2}, 1) - \frac{\text{li}(v)}{p_{k-l+2}} \right) \\ & \ll \frac{v \log y}{\log v} + \text{li}(v)(\log \log v)^{l-2}. \end{aligned}$$

*Proof.* Let  $E(t; r, 1) = \pi(t; r, 1) - \frac{\text{li}(t)}{r-1}$ . Then we have

$$\begin{aligned} & \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} \left( \pi(v, p_{k-l+2}, 1) - \frac{\text{li}(v)}{p_{k-l+2} - 1} \right) \\ & = \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} E(v; p_{k-l+2}, 1) \\ & \ll \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} |E(v; p_{k-l+2}, 1)|. \end{aligned}$$

Let  $\Omega(m)$  denote the number of divisors of  $m$  which are primes or prime powers. We use the estimate  $\Omega(m) \ll \log m$  to get

$$\begin{aligned} & \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} |E(v; p_{k-l+2}, 1)| \\ & \leq \log(y^k) \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} |E(v; p_{k-l+2}, 1)| \sum_{\substack{p_{k-l+3} | p_{k-l+2} - 1 \\ p_3 \leq v^{1/9}}} \sum_{\substack{p_{k-l+4} | p_{k-l+3} - 1 \\ p_{k-l+4} \leq v^{1/27}}} \cdots \sum_{q \leq y^k} \sum_{\substack{a \in \mathbb{N} \\ q^a | p_{k-1}}} 1 \\ & \leq \log(y^k) \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} |E(v; p_{k-l+2}, 1)| \sum_{\substack{p_{k-l+3} | p_{k-l+2} - 1 \\ p_3 \leq v^{1/9}}} \sum_{\substack{p_{k-l+4} | p_{k-l+3} - 1 \\ p_{k-l+4} \leq v^{1/27}}} \cdots \sum_{\substack{p_k \leq v^{1/3^{k-1}} \\ p_k | p_{k-1} - 1}} \Omega(p_k - 1) \\ & \ll \log y \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} |E(v; p_{k-l+2}, 1)| \sum_{\substack{p_{k-l+3} | p_{k-l+2} - 1 \\ p_3 \leq v^{1/9}}} \sum_{\substack{p_{k-l+4} | p_{k-l+3} - 1 \\ p_{k-l+4} \leq v^{1/27}}} \cdots \sum_{\substack{p_k \leq v^{1/3^{k-1}} \\ p_k | p_{k-1} - 1}} \log t. \end{aligned}$$

Continuing in this manner we obtain

$$\begin{aligned} & \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} |E(v; p_{k-l+2}, 1)| \\ & \ll \log y (\log v)^{l-1} \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} |E(v; p_{k-l+2}, 1)| \ll \frac{v \log y}{\log t} \end{aligned}$$

using Bombieri–Vinogradov. As for the difference between

$$\sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} \frac{\text{li}(v)}{p_{k-l+2} - 1}$$

and

$$(30) \quad \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} \frac{\text{li}(v)}{p_{k-l+2}}$$

we get that it is

$$\begin{aligned} & \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} \frac{\text{li}(v)}{p_{k-l+2}(p_{k-l+2} - 1)} \\ & \leq \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{i=1}^{\infty} \frac{\text{li}(v)}{(ip_{k-l+3} + 1)(ip_{k-l+3})} \\ & \ll \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+3} \in \mathcal{P}_{p_{k-l+4}} \\ p_{k-l+3} \leq v^{1/9}}} \frac{\text{li}(v)}{p_{k-l+3}^2} \\ & \ll \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+3} \in \mathcal{P}_{p_{k-l+4}} \\ p_{k-l+3} \leq v^{1/9}}} \frac{\text{li}(v)}{p_{k-l+3} q^a} \\ & \ll \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \frac{\text{li}(v)(\log \log v)^{l-2}}{q^{2a}} \\ & \ll \sum_{q \leq y^k} \frac{\text{li}(v)(\log \log v)^{l-2} \log q}{q^2} \\ & \ll \text{li}(v)(\log \log v)^{l-2} \end{aligned}$$

using the Brun–Titchmarsh inequality (11), the inequality  $p_{k-l+3} \geq q^a$  and noting that the sum over  $q$  converges.  $\square$

**Lemma 23.** *For all  $x > e^{e^e}$  and  $t > e^e$ ,*

$$\begin{aligned} \sum_{p \leq t} h_k(p) &= \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq t^{1/3^{k-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq t^{1/3^{k-2}}}} \cdots \sum_{\substack{p_2 \in \mathcal{P}_{p_3} \\ p_2 \leq t^{1/3}}} \pi(t; p_2, 1) \\ &\quad + O\left(t^{1-1/3^k} \log t (\log \log t)^{k-2} y^k + \frac{t(\log \log t)^{k-2} \log y}{\log t}\right). \end{aligned}$$

*Proof.* For a prime  $p$ ,

$$\begin{aligned} h_k(p) &= \sum_{p_1|p} \sum_{p_2|p_1-1} \cdots \sum_{p_k|p_{k-1}-1} \sum_{q \leq y^k} \nu_q(p_k - 1) \log q \\ &= \sum_{p_2|p-1} \cdots \sum_{p_k|p_{k-1}-1} \sum_{q \leq y^k} \nu_q(p_k - 1) \log q \end{aligned}$$

since the only prime which can divide  $p$  is  $p$  itself. Hence

$$\begin{aligned} \sum_{p \leq t} h_k(p) &= \sum_{p \leq t} \sum_{p_2|p-1} \cdots \sum_{p_k|p_{k-1}-1} \sum_{q \leq y^k} \nu_q(p_k - 1) \log q \\ &= \sum_{p \leq t} \sum_{p_2|p-1} \cdots \sum_{p_k|p_{k-1}-1} \sum_{q \leq y^k} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ a \in \mathbb{N}}} \log q \\ &= \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{p_k \in \mathcal{P}_{q^a}} \sum_{p_{k-1} \in \mathcal{P}_{p_k}} \cdots \sum_{\substack{p_2 \in \mathcal{P}_{p_3} \\ p \in \overline{\mathcal{P}}_{p_2}}} \sum_{p \leq t} 1 \\ &= \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{p_k \in \mathcal{P}_{q^a}} \sum_{p_{k-1} \in \mathcal{P}_{p_k}} \cdots \sum_{p_2 \in \mathcal{P}_{p_3}} \pi(t; p_2, 1). \end{aligned}$$

We wish to approximate  $\pi(t; p_2, 1)$  by  $\frac{\text{li}(t)}{p_2-1}$  and use the Bombieri-Vinogradov Theorem to deal with the error. However this approximation only allows primes up to say  $t^{1/3}$ . So we use the estimations in Lemma 20 to bound these errors. We will see that the main contribution comes from  $p_i \leq t^{1/3^{i-1}}$  and  $q^a \leq t^{1/3^k}$ .

Using Lemma 20, we get for large  $q^a$

$$\sum_{q \leq y^k} \log q \sum_{\substack{a \in \mathbb{N} \\ q^a > t^{1/3^k}}} \sum_{p_k \in \mathcal{P}_{q^a}} \sum_{p_{k-1} \in \mathcal{P}_{p_k}} \cdots \sum_{p_3 \in \mathcal{P}_{p_2}} \pi(t; p_2, 1) \ll \sum_{q \leq y^k} \log q \sum_{\substack{a \in \mathbb{N} \\ q^a > t^{1/3^k}}} \frac{t \log t (\log \log t)^{k-2}}{q^a}.$$

By geometric estimates, if  $a^*$  is the smallest  $a$  where  $q^a > t^{1/3^k}$ , then we get that the above is

$$\begin{aligned} &\ll t \log t (\log \log t)^{k-2} \sum_{q \leq y^k} \frac{\log q}{q^{a^*}} \\ &\leq t^{1-1/3^k} \log t (\log \log t)^{k-2} \sum_{q \leq y^k} \log q \\ &\ll t^{1-1/3^k} \log t (\log \log t)^{k-2} y^k. \end{aligned}$$

Now suppose  $q^a \leq t^{1/3^k}$ . Let  $l$  be the last index (supposing one exists) where  $p_i > t^{1/3^{i-1}}$ . By using (24) where  $l$  ranges from 2 to  $k$ , we can bound the large values of the  $p_i$ .

$$\begin{aligned}
& \sum_{q \leq y^k} \log q \sum_{\substack{a \in \mathbb{N} \\ q^a \leq t^{1/3^k}}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq t^{1/3^{k-1}}}} \cdots \sum_{\substack{p_{l+1} \in \mathcal{P}_{p_{l+2}} \\ p_{l+1} \leq t^{1/3^l}}} \sum_{\substack{p_l \in \mathcal{P}_{p_{l+1}} \\ p_l > t^{1/3^{l-1}}}} \sum_{p_{l-1} \in \mathcal{P}_{p_l}} \cdots \sum_{p_2 \in \mathcal{P}_{p_3}} \pi(t; p_2, 1) \\
& \ll \sum_{q \leq y^k} \log q \sum_{\substack{a \in \mathbb{N} \\ q^a \leq t^{1/3^k}}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq t^{1/3^{k-1}}}} \cdots \sum_{\substack{p_{l+2} \in \mathcal{P}_{p_{l+3}} \\ p_{l+2} \leq t^{1/3^{l+1}}}} \sum_{\substack{p_{l+1} \in \mathcal{P}_{p_{l+2}} \\ p_{l+1} > t^{1/3^l}}} \frac{(p_l)^{l-1} t}{\phi(p_l)^l \log t} \\
& \ll \sum_{q \leq y^k} \log q \sum_{\substack{a \in \mathbb{N} \\ q^a \leq t^{1/3^k}}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq t^{1/3^{k-1}}}} \cdots \sum_{\substack{p_{l+2} \in \mathcal{P}_{p_{l+3}} \\ p_{l+2} \leq t^{1/3^{l+1}}}} \sum_{\substack{p_{l+1} \in \mathcal{P}_{p_{l+2}} \\ p_{l+1} > t^{1/3^l}}} \frac{t}{p_{l+1} \log t}
\end{aligned}$$

since  $p_l$  is prime and  $l \leq k$ . By Brun-Titchmarsh (11) we get

$$\begin{aligned}
& \ll \sum_{q \leq y^k} \log q \sum_{\substack{a \in \mathbb{N} \\ q^a \leq t^{1/3^k}}} \frac{t(\log \log t)^{k-l}}{q^a \log t} \\
& \ll \sum_{q \leq y^k} \frac{t(\log \log t)^{k-l} \log q}{q \log t} \\
& \ll \frac{t(\log \log t)^{k-2} \log y}{\log t}
\end{aligned}$$

by (3) and since  $l \geq 2$ . Hence we get

$$\begin{aligned}
\sum_{p \leq t} h_k(p) &= \sum_{q \leq y^k} \log q \sum_{\substack{a \in \mathbb{N} \\ q^a \leq t^{1/3^k}}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq t^{1/3^{k-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq t^{1/3^{k-2}}}} \cdots \sum_{\substack{p_2 \in \mathcal{P}_{p_3} \\ p_2 \leq t^{1/3}}} \pi(t, p_2, 1) \\
&+ O\left(t^{1-1/3^k} \log t (\log \log t)^{k-2} y^k + \frac{t(\log \log t)^{k-2} \log y}{\log t}\right)
\end{aligned}$$

finishing the lemma.  $\square$

## 12. EVALUATION OF THE MAIN TERM

Now we'll deal with the main term from Lemma 23. We will deal with estimating the individual sums recursively. Hence we wish to make the following definition.

**Definition 24.** Let  $2 \leq l \leq k$  and  $2 \leq u \leq t$ . Then define

$$g_{k,l}(u) = \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq u^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq u^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq u^{1/3}}} \pi(u; p_{k-l+2}, 1).$$

Note that  $g_{k,k}(t)$  is the summation in Lemma 23. Next we'll exhibit the recursive formula satisfied by the  $g_{k,l}$ .

**Lemma 25.** *Let  $3 \leq l \leq k$ , then*

$$(31) \quad g_{k,l}(v) = \text{li}(v) \int_2^{v^{1/3}} \frac{1}{u^2} g_{k,l-1}(u) du + O\left(\frac{v(\log \log v)^{l-2} \log y}{\log v}\right).$$

*Proof.* We'll proceed by approximating  $\pi$  by  $\text{li}$  and then use partial summation to recover  $\pi$ . Using Lemma 22 we get

$$\begin{aligned} g_{k,l}(v) &= \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} \pi(v; p_{k-l+2}, 1) \\ &\ll \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} \frac{\text{li}(v)}{p_{k-l+2}} + O\left(\frac{v \log y}{\log v} + \text{li}(v)(\log \log v)^{l-2}\right). \end{aligned}$$

We use Euler summation on the inner sum to get

$$\sum_{\substack{p_{k-l+2} \in \mathcal{P}_{p_{k-l+3}} \\ p_{k-l+2} \leq v^{1/3}}} \frac{1}{p_{k-l+2}} = \frac{\pi(v^{1/3}; p_{k-l+3}, 1)}{v^{1/3}} + \int_2^{v^{1/3}} \frac{\pi(u; p_{k-l+3}, 1)}{u^2} du$$

and so we get that

$$\begin{aligned} g_{k,l}(v) &= \text{li}(v) \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+3} \in \mathcal{P}_{p_{k-l+4}} \\ p_{k-l+3} \leq v^{1/3}}} \left( \frac{\pi(v^{1/3}; p_{k-l+3}, 1)}{v^{1/3}} \right. \\ &\quad \left. + \int_2^{v^{1/3}} \frac{\pi(u; p_{k-l+3}, 1)}{u^2} du \right) + O\left(\frac{v \log y}{\log v} + \text{li}(v)(\log \log v)^{l-2}\right). \end{aligned}$$

Inside the sum by trivially estimating  $\pi(x; q, 1)$  by  $x/q$  inside the sum and using Brun–Titchmarsh (11) we get

$$\begin{aligned} &\sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+3} \in \mathcal{P}_{p_{k-l+4}} \\ p_{k-l+3} \leq v^{1/3}}} \frac{\pi(v^{1/3}; p_{k-l+3}, 1)}{v^{1/3}} \\ &\ll \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq v^{1/3^{l-1}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ p_{k-1} \leq v^{1/3^{l-2}}}} \cdots \sum_{\substack{p_{k-l+3} \in \mathcal{P}_{p_{k-l+4}} \\ p_{k-l+3} \leq v^{1/3}}} \frac{1}{p_{k-l+3}} \\ &\ll \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \frac{(\log \log v)^{l-2}}{q^a} \\ &\ll \sum_{q \leq y^k} \log q \frac{(\log \log v)^{l-2}}{q} \\ &\ll (\log \log v)^{l-2} \log y. \end{aligned}$$

Multiplying through by  $\text{li}(v)$  finishes the lemma. □

We now require a lemma to find the asymptotic formula for  $h_k$  using the previous recurrence relation

**Lemma 26.** *Let  $2 \leq l \leq k$ .*

$$g_{k,l}(u) = \frac{ku(\log \log u)^{l-1} \log y}{(l-1)! \log u} + O\left(\frac{u(\log \log u)^{l-1}}{\log u} + \frac{u(\log \log u)^{l-2} \log^2 y}{\log u}\right)$$

which implies

$$\begin{aligned} \sum_{p \leq t} h_k(p) &= \frac{kt(\log \log t)^{k-1} \log y}{(k-1)! \log t} + O\left(\frac{t(\log \log t)^{k-1}}{\log t} \right. \\ &\quad \left. + \frac{t(\log \log t)^{k-2} \log^2 y}{\log t} + t^{1-1/3^k} \log t (\log \log t)^{k-2} y^k\right). \end{aligned}$$

*Proof.* The second formula is derived from the first by setting  $l = k$ ,  $u = t$  and using Lemma 23. We'll proceed with the first formula by induction on  $l$ . Using the estimates we obtained via Bombieri–Vinogradov in Lemma 22, we have for  $l = 2$

$$\begin{aligned} g_{k,2}(u) &= \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq u^{1/3}}} \pi(u; p_k, 1) \\ &= \text{li}(u) \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q^a} \\ p_k \leq u^{1/3}}} \frac{1}{p_k} + O\left(\text{li}(u) + \frac{u \log y}{\log u}\right). \end{aligned}$$

We then use (13) and

$$\log \log(u^{1/3}) = \log \log u + O(1)$$

to get

$$\begin{aligned} g_{k,2}(u) &= \text{li}(u) \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \left( \frac{\log \log u^{1/3}}{\phi(q^a)} + O\left(\frac{\log(q^a)}{\phi(q^a)}\right) \right) + O\left(\frac{u \log y}{\log u}\right) \\ &= \text{li}(u)(\log \log u + O(1)) \sum_{q \leq y^k} \log q \sum_{a \in \mathbb{N}} \left( \frac{1}{q^a} + O\left(\frac{1}{q^{a+1}}\right) \right) + O\left(\text{li}(u) \sum_{q \leq y^k} \log^2 q \sum_{a \in \mathbb{N}} \frac{a}{q^a}\right) \\ &\quad + O\left(\frac{u \log y}{\log u}\right) \\ &= \text{li}(u)(\log \log u + O(1)) \sum_{q \leq y^k} \left( \frac{\log q}{q} + O\left(\frac{\log q}{q^2}\right) \right) + O\left(\text{li}(u) \sum_{q \leq y^k} \frac{\log^2 q}{q} + \frac{u \log y}{\log u}\right) \\ &= \text{li}(u) \log \log u \log(y^k) + O\left(\text{li}(u)(\log y + \log \log u + \log^2 y) + \frac{u \log y}{\log u}\right) \\ &= \frac{ku \log \log u \log y}{\log u} + O\left(\frac{u \log \log u}{\log u} + \frac{u \log^2 y}{\log u}\right), \end{aligned}$$



completing the base case. Now using Lemma 25 we get

$$\begin{aligned}
g_{k,l}(v) &= \text{li}(v) \int_2^{v^{1/3}} \frac{1}{u^2} g_{k,l-1}(u) du + O\left(\frac{v(\log \log v)^{l-2} \log y}{\log v}\right) \\
&= \text{li}(v) \int_2^{v^{1/3}} \frac{1}{u^2} \left( \frac{ku(\log \log u)^{l-2} \log y}{(l-2)! \log u} + O\left(\frac{u(\log \log u)^{l-2}}{\log u} + \frac{u(\log \log u)^{l-3} \log^2 y}{\log u}\right) \right) du + O\left(\frac{v(\log \log v)^{l-2} \log y}{\log v}\right) \\
&= \text{li}(v) \int_2^{v^{1/3}} \left( \frac{k(\log \log u)^{l-2} \log y}{(l-2)! u \log u} + O\left(\frac{(\log \log u)^{l-2}}{u \log u} + \frac{(\log \log u)^{l-3} \log^2 y}{u \log u}\right) \right) du \\
&\quad + O\left(\frac{v(\log \log v)^{l-2} \log y}{\log v}\right) \\
&= \frac{k \text{li}(v)(\log \log v^{1/3})^{l-1} \log y}{(l-1)!} + O\left(\text{li}(v)(\log \log v^{1/3})^{l-1} + \text{li}(v)(\log \log v^{1/3})^{l-2} \log^2 y + \frac{v(\log \log v)^{l-2} \log y}{\log v}\right).
\end{aligned}$$

Once again by using

$$\log \log v^{1/3} = \log \log v + O(1)$$

we get

$$\begin{aligned}
&\frac{kv(\log \log v)^{l-1} \log y}{(l-1)! \log v} + O\left(\frac{v(\log \log v)^{l-1}}{\log v} + \frac{v(\log \log v)^{l-2} \log^2 y}{\log v} + \frac{v(\log \log v)^{l-2} \log y}{\log v}\right) \\
&= \frac{kv(\log \log v)^{l-1} \log y}{(l-1)! \log v} + O\left(\frac{v(\log \log v)^{l-1}}{\log v} + \frac{v(\log \log v)^{l-2} \log^2 y}{\log v}\right),
\end{aligned}$$

completing the induction. □

### 13. THE PROOF OF THE FIRST MOMENT

We now are in a position to prove the proposition for the first moment.

*Proof of Proposition 14.*

$$\begin{aligned}
M_1(x) &= \sum_{p \leq x} \frac{h_k(p)}{p} \\
&= \sum_{p \leq e^e} \frac{h_k(p)}{p} + \sum_{e^e < p \leq x} \frac{h_k(p)}{p} \\
&= O(1) + \sum_{e^e < p \leq x} h_k(p) \left( \frac{1}{x} + \int_p^x \frac{dt}{t^2} \right) \\
&= O(1) + \frac{1}{x} \sum_{e^e < p \leq x} h_k(p) + \int_{e^e}^x \frac{dt}{t^2} \sum_{e^e < p \leq t} h_k(p).
\end{aligned}$$

Using  $t = x$  in Lemma 26 we get that

$$\sum_{e^e < p \leq x} h_k(p) \ll \frac{xy^{k-1} \log y}{\log x}$$

and since

$$\sum_{e^e < p \leq t} h_k(p)$$

differs from

$$\sum_{p \leq t} h_k(p)$$

by a constant, we get that

$$\begin{aligned} M_1(x) = O(1) + \frac{1}{x} O\left(\frac{xy^{k-1} \log y}{\log x}\right) + \int_{e^e}^x \frac{dt}{t^2} \left( \frac{kt(\log \log t)^{k-1} \log y}{(k-1)! \log t} + O\left(\frac{t(\log \log t)^{k-1}}{\log t} \right. \right. \\ \left. \left. + \frac{t(\log \log t)^{k-2} \log^2 y}{\log t} + t^{1-1/3^k} \log t (\log \log t)^{k-2} y^k \right) \right) \end{aligned}$$

using Lemma 26. Noting that

$$\begin{aligned} \int_{e^e}^x \frac{dt}{t^2} t^{1-1/3^k} \log t (\log \log t)^{k-2} y^k \\ = \int_{e^e}^x \frac{y^k dt}{t^{1+\epsilon}} \\ \ll y^k \end{aligned}$$

yields

$$\begin{aligned} O(y^k) + O\left(\frac{y^{k-1} \log y}{\log x}\right) + \int_{e^e}^x \frac{dt}{t^2} \left( \frac{kt(\log \log t)^{k-1} \log y}{(k-1)! \log t} + O\left(\frac{t(\log \log t)^{k-1}}{\log t} \right. \right. \\ \left. \left. + \frac{t(\log \log t)^{k-2} \log^2 y}{\log t} \right) \right) \\ = O(y^k) + \frac{k(\log \log x)^k \log y}{k(k-1)!} + O\left((\log \log x)^k + (\log \log x)^{k-1} \log^2 y\right) \\ = \frac{y^k \log y}{(k-1)!} + O(y^k) \end{aligned}$$

as needed. □

#### 14. THE PROOF OF THE SECOND MOMENT

We now turn our attention to the second moment. Our first lemma will bound the case where  $p_3 = r_3$  and then we'll use the summations from Lemma 21 to take care of the rest.

**Lemma 27.**

$$\begin{aligned} & \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{a_1, a_2 \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q_1^{a_1}} \\ r_k \in \mathcal{P}_{q_2^{a_2}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ r_{k-1} \in \mathcal{P}_{r_k}}} \cdots \sum_{\substack{p_3 \in \mathcal{P}_{p_4} \\ r_3 \in \mathcal{P}_{r_4}}} \sum_{\substack{s \in \mathcal{P}_{p_3} \cap \mathcal{P}_{r_3} \\ p \leq t \\ p \in \mathcal{P}_s}} 1 \\ & \ll t^{1-\epsilon} y^k \log y + \frac{t(\log \log t)^{2k-2}}{\log t} \log^2 y \end{aligned}$$

for some  $\epsilon > 0$ .

*Proof.* Our sum is

$$\begin{aligned} & \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{a_1, a_2 \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q_1^{a_1}} \\ r_k \in \mathcal{P}_{q_2^{a_2}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ r_{k-1} \in \mathcal{P}_{r_k}}} \cdots \sum_{\substack{p_3 \in \mathcal{P}_{p_4} \\ r_3 \in \mathcal{P}_{r_4}}} \sum_{\substack{s \in \mathcal{P}_{p_3} \cap \mathcal{P}_{r_3} \\ p \leq t \\ p \in \mathcal{P}_s}} 1 \\ & = \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{a_1, a_2 \in \mathbb{N}} \sum_{\substack{p_k \in \mathcal{P}_{q_1^{a_1}} \\ r_k \in \mathcal{P}_{q_2^{a_2}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ r_{k-1} \in \mathcal{P}_{r_k}}} \cdots \sum_{\substack{p_3 \in \mathcal{P}_{p_4} \\ r_3 \in \mathcal{P}_{r_4}}} \sum_{s \in \mathcal{P}_{p_3 r_3}} \pi(t; s, 1). \end{aligned}$$

We split up into two cases. If  $q_1^{a_1} q_2^{a_2} > t^\alpha$ , then suppose  $q_1^{a_1} > t^{\alpha/2}$ . (the other case is analogous) We get from the trivial bound on  $\pi(t; s, 1)$  that

$$\begin{aligned} & \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{\substack{a_1, a_2 \in \mathbb{N} \\ q_1^{a_1} > t^{\frac{\alpha}{2}}}} \sum_{\substack{p_k \in \mathcal{P}_{q_1^{a_1}} \\ r_k \in \mathcal{P}_{q_2^{a_2}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ r_{k-1} \in \mathcal{P}_{r_k}}} \cdots \sum_{\substack{p_3 \in \mathcal{P}_{p_4} \\ r_3 \in \mathcal{P}_{r_4}}} \sum_{s \in \mathcal{P}_{p_3 r_3}} \pi(t; s, 1) \\ & = \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{\substack{a_1, a_2 \in \mathbb{N} \\ q_1^{a_1} > t^{\frac{\alpha}{2}}}} \sum_{\substack{p_k \in \mathcal{P}_{q_1^{a_1}} \\ r_k \in \mathcal{P}_{q_2^{a_2}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ r_{k-1} \in \mathcal{P}_{r_k}}} \cdots \sum_{\substack{p_3 \in \mathcal{P}_{p_4} \\ r_3 \in \mathcal{P}_{r_4}}} \sum_{s \in \mathcal{P}_{p_3 r_3}} \frac{t \log t}{s} \\ & = \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{\substack{a_1, a_2 \in \mathbb{N} \\ q_1^{a_1} > t^{\frac{\alpha}{2}}}} \sum_{\substack{p_k \in \mathcal{P}_{q_1^{a_1}} \\ r_k \in \mathcal{P}_{q_2^{a_2}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ r_{k-1} \in \mathcal{P}_{r_k}}} \cdots \sum_{\substack{p_3 \in \mathcal{P}_{p_4} \\ r_3 \in \mathcal{P}_{r_4}}} \frac{t \log t \log \log t}{p_3 r_3} \\ & = \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{\substack{a_1, a_2 \in \mathbb{N} \\ q_1^{a_1} > t^{\frac{\alpha}{2}}}} \frac{t \log t (\log \log t)^{2k-3}}{q_1^{a_1} q_2^{a_2}}. \end{aligned}$$

By letting  $A = \min\{a | q_1^a > t^{\frac{\alpha}{2}}\}$  we get

$$\begin{aligned} & \ll \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \frac{t \log t (\log \log t)^{k-1}}{q_1^A q_2} \\ & \leq t^{1-\frac{\alpha}{2}} \log t (\log \log t)^{2k-3} \sum_{q_1 \leq y^k} \log q_1 \sum_{q_2 \leq y^k} \frac{\log q_2}{q} \\ & \ll t^{1-\epsilon} y^k \log y. \end{aligned}$$

If  $q_1^{a_1} q_2^{a_2} > t^\alpha$ , then by Lemma 21 part (d) we get

$$\begin{aligned}
& \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{\substack{a_1, a_2 \in \mathbb{N} \\ q_1^{a_1} q_2^{a_2} \leq t^\alpha}} \sum_{\substack{p_k \in \mathcal{P}_{q_1^{a_1}} \\ r_k \in \mathcal{P}_{q_2^{a_2}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ r_{k-1} \in \mathcal{P}_{r_k}}} \cdots \sum_{\substack{p_3 \in \mathcal{P}_{p_4} \\ r_3 \in \mathcal{P}_{r_4}}} \sum_{s \in \mathcal{P}_{p_3 r_3}} \pi(t; s, 1) \\
& \ll \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{\substack{a_1, a_2 \in \mathbb{N} \\ q_1^{a_1} q_2^{a_2} \leq t^\alpha}} \frac{t(\log \log t)^{2k-2}}{q_1^{a_1} q_2^{a_2} \log t} \\
& \ll \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \frac{t(\log \log t)^{2k-2}}{q_1 q_2 \log t} \\
& = \frac{t(\log \log t)^{2k-2}}{\log t} \left( \sum_{q \leq y^k} \frac{\log q}{q} \right)^2 \\
& \ll \frac{t(\log \log t)^{2k-2}}{\log t} \log^2 y
\end{aligned}$$

by (3), completing the lemma.  $\square$

We now have enough to finish the second moment which is the final piece of the puzzle.

*Proof of Proposition 15.*

$$\begin{aligned}
\sum_{p \leq t} h_k(p)^2 &= \sum_{p \leq x} \left( \sum_{p_1 | p} \sum_{p_2 | p_1 - 1} \cdots \sum_{p_k | p_{k-1} - 1} \sum_{q \leq y^k} \nu_q(p_k - 1) \log q \right)^2 \\
&= \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{\substack{a_1, a_2 \in \mathbb{N} \\ q_1^{a_1} q_2^{a_2} \leq t^\alpha}} \sum_{\substack{p_k \in \mathcal{P}_{q_1^{a_1}} \\ r_k \in \mathcal{P}_{q_2^{a_2}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ r_{k-1} \in \mathcal{P}_{r_k}}} \cdots \sum_{\substack{p_2 \in \mathcal{P}_{p_3} \\ r_2 \in \mathcal{P}_{r_3}}} \sum_{\substack{p \leq t \\ p \in \mathcal{P}_{p_2} \\ p \in \mathcal{P}_{r_2}}} 1
\end{aligned}$$

since the condition  $p_1 \mid p$  only occurs if  $p_1 = p$ . We then split up the sum according to whether or not  $p_2 = r_2$ . Lemma 27 deals with the part where  $s = p_2 = r_2$  leaving us with

$$\begin{aligned}
& \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{\substack{a_1, a_2 \in \mathbb{N} \\ q_1^{a_1} q_2^{a_2} \leq t^\alpha}} \sum_{\substack{p_k \in \mathcal{P}_{q_1^{a_1}} \\ r_k \in \mathcal{P}_{q_2^{a_2}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ r_{k-1} \in \mathcal{P}_{r_k}}} \cdots \sum_{\substack{p_2 \in \mathcal{P}_{p_3} \\ r_2 \in \mathcal{P}_{r_3} \\ p_2 \neq r_2}} \sum_{\substack{p \leq t \\ p \in \mathcal{P}_{p_2} \\ p \in \mathcal{P}_{r_2}}} 1 \\
& + O\left(t^{1-\epsilon} y^k \log y + \frac{t(\log \log t)^{2k-2}}{\log t} \log^2 y\right).
\end{aligned}$$

The sum becomes

$$\sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{\substack{a_1, a_2 \in \mathbb{N} \\ q_1^{a_1} q_2^{a_2} \leq t^\alpha}} \sum_{\substack{p_k \in \mathcal{P}_{q_1^{a_1}} \\ r_k \in \mathcal{P}_{q_2^{a_2}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ r_{k-1} \in \mathcal{P}_{r_k}}} \cdots \sum_{\substack{p_2 \in \mathcal{P}_{p_3} \\ r_2 \in \mathcal{P}_{r_3}}} \pi(t; p_2 r_2, 1).$$

If  $q_1^{a_1} > t^{\alpha_1}$ , then so is  $p_2$ , and hence by (26) we get

$$\begin{aligned}
& \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{\substack{a_1, a_2 \in \mathbb{N} \\ q_1^{a_1} > t^{\alpha_1}}} \sum_{\substack{p_k \in \mathcal{P}_{q_1^{a_1}} \\ r_k \in \mathcal{P}_{q_2^{a_2}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ r_{k-1} \in \mathcal{P}_{r_k}}} \cdots \sum_{\substack{p_3 \in \mathcal{P}_{p_4} \\ r_3 \in \mathcal{P}_{r_4}}} \frac{t \log^2 t}{p_3 r_3} \\
& \ll \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{\substack{a_1, a_2 \in \mathbb{N} \\ q_1^{a_1} > t^{\alpha_1}}} \frac{t \log^2 t (\log \log t)^{2k-4}}{q_1^{a_1} q_2^{a_2}} \\
& \ll t^{1-\alpha_1} \log^2 t (\log \log t)^{2k-4} \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{a_2 \in \mathbb{N}} \frac{1}{q_2^{a_2}} \\
& \ll t^{1-\alpha_1} \log^2 t (\log \log t)^{2k-4} \sum_{q_1, q_2 \leq y^k} \frac{\log q_1 \log q_2}{q_2} \\
& \ll t^{1-\alpha_1} \log^2 t (\log \log t)^{2k-4} (y^k \log y).
\end{aligned}$$

We similarly get the same bound if  $q_2^{a_2} > t^{\alpha_1}$ . If neither of  $q_1^{a_1}, q_2^{a_2}$  exceed  $t^{\alpha_1}$ , then by (28) and using that for  $b_i = q_i^{a_i}$

$$\frac{b_i}{\phi(b_i)} \ll 1, \frac{1}{\phi(b_i)} \ll \frac{1}{b_i},$$

we get

$$\begin{aligned}
& \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{\substack{a_1, a_2 \in \mathbb{N} \\ q_1^{a_1}, q_2^{a_2} \leq t^{\alpha_1}}} \sum_{\substack{p_k \in \mathcal{P}_{q_1^{a_1}} \\ r_k \in \mathcal{P}_{q_2^{a_2}}}} \sum_{\substack{p_{k-1} \in \mathcal{P}_{p_k} \\ r_{k-1} \in \mathcal{P}_{r_k}}} \cdots \sum_{\substack{p_i \in \mathcal{P}_{p_{i+1}} \\ r_i \in \mathcal{P}_{r_{i+1}}}} \sum_{\substack{p_{i-1} \in \mathcal{P}_{p_i} \\ r_{i-1} \in \mathcal{P}_{r_i}}} \cdots \sum_{\substack{p_2 \in \mathcal{P}_{p_3} \\ r_2 \in \mathcal{P}_{r_3}}} \pi(t; p_2 r_2, 1) \\
& \ll \sum_{q_1, q_2 \leq y^k} \log q_1 \log q_2 \sum_{\substack{a_1, a_2 \in \mathbb{N} \\ q_1^{a_1}, q_2^{a_2} \leq t^{\alpha_1}}} \frac{t (\log \log t)^{2k-2}}{q_1^{a_1} q_2^{a_2} \log t} \\
& \ll \frac{t (\log \log t)^{2k-2}}{\log t} \sum_{q_1, q_2 \leq y^k} \frac{\log q_1 \log q_2}{q_1 q_2} \\
& \ll \frac{t (\log \log t)^{2k-2} \log^2 y}{\log t}.
\end{aligned}$$

Hence the above gives us that

$$\sum_{p \leq t} h_k(p)^2 \ll t^{1-\epsilon} y^k \log y + \frac{t (\log \log t)^{2k-2} \log^2 y}{\log t}.$$

Using partial summation we have

$$\begin{aligned}
M_2(x) &= \sum_{p \leq x} \frac{h_k(p)^2}{p} = \sum_{p \leq e^e} \frac{h_k(p)^2}{p} + \frac{1}{x} \sum_{e^e \leq p \leq x} h_k(p)^2 + \int_{e^e}^x \frac{dt}{t^2} \sum_{e^e \leq p \leq t} h_k(p)^2 \\
&\ll 1 + \frac{1}{x} \left( x^{1-\epsilon} y^k \log y + \frac{x(\log \log x)^{2k-2} \log^2 y}{\log x} \right) \\
&\quad + \int_{e^e}^x \left( t^{-1-\epsilon} y^k \log y + \frac{(\log \log t)^{2k-2} \log^2 y}{t \log t} \right) dt \\
&\ll \frac{y^{2k-2} \log^2 y}{\log x} + x^{-\epsilon} y^k \log y + (\log \log x)^{2k-1} \log^2 y \\
&\ll y^{2k-1} \log^2 y
\end{aligned}$$

completing the proof of Proposition 15 and hence Theorem 1.  $\square$

## 15. THEOREM 2

We now turn our attention to the proof of Theorem 2. It will be necessary to use the following upper bound for the Carmichael function of a product.

**Lemma 28.** *Let  $a, b$  be natural numbers, then*

$$(32) \quad \lambda(ab) \leq b\lambda(a).$$

*Proof.* We first note that it suffices to show the inequality whenever  $b$  is prime, because if

$$b = p_1 \dots p_k$$

where the  $p_i$  are not necessarily distinct, then repeated use of the theorem where  $b$  is prime yields

$$\lambda(ab) = \lambda(ap_1 \dots p_k) \leq p_1 \lambda(ap_2 \dots p_k) \leq \dots \leq p_1 \dots p_k \lambda(a) = b\lambda(a).$$

If  $b$  is a prime which divides  $a$ , then

$$a = b^e p_1^{e_1} \dots p_k^{e_k} \text{ and } ab = b^{e+1} p_1^{e_1} \dots p_k^{e_k}.$$

Therefore

$$\begin{aligned}
\lambda(ab) &= \text{lcm} \left( \lambda(b^{e+1}), \lambda(p_1^{e_1}), \dots, \lambda(p_k^{e_k}) \right) \\
&\leq \text{lcm} \left( b\lambda(b^e), \lambda(p_1^{e_1}), \dots, \lambda(p_k^{e_k}) \right) \\
&\leq b * \text{lcm} \left( \lambda(b^e), \lambda(p_1^{e_1}), \dots, \lambda(p_k^{e_k}) \right) \\
&= b\lambda(a)
\end{aligned}$$

where the first inequality is in fact an equality if  $b^e = 4$ . Also note that in this case, it would not be hard to show that  $\lambda(ab) \mid b\lambda(a)$ . If  $(a, b) = 1$ , then

$$\begin{aligned}
\lambda(ab) &= \text{lcm}\left(b-1, \lambda(p_1^{e_1}), \dots, \lambda(p_k^{e_k})\right) \\
&\leq (b-1) \text{lcm}\left(\lambda(p_1^{e_1}), \dots, \lambda(p_k^{e_k})\right) \\
&< b\lambda(a),
\end{aligned}$$

ending the proposition.  $\square$

Suppose that  $g(n)$  is an arithmetic function of the form  $\phi(h(n))$  where  $h(n)$  is a  $(k-1)$ -fold iterate involving  $\phi$  and  $\lambda$ . Then we can use equation (32) to get

$$\lambda_{l+k}(n) \leq \lambda_l(g(n)) \leq \lambda_l\left(\frac{g(n)}{\lambda_k(n)}\lambda_k(n)\right) \leq \lambda_{l+k}(n)\frac{g(n)}{\lambda_k(n)}.$$

Since  $g(n) \leq n$  we have that

$$\frac{g(n)}{\lambda_k(n)} \leq \frac{n}{\lambda_k(n)} = \exp\left(\frac{1}{(k-1)!}(1+o_k(1))(\log \log n)^k \log \log \log n\right)$$

by Theorem 1 and hence

$$\lambda_{l+k}(n) \leq \lambda_l(g(n)) \leq \lambda_l\left(\frac{g(n)}{\lambda_k(n)}\lambda_k(n)\right) \leq \lambda_{l+k}(n) \exp\left(\frac{1}{(k-1)!}(\log \log n)^k(1+o_k(1)) \log \log \log n\right).$$

From the fact that

$$\lambda_{l+k}(n) = n \exp\left(-\frac{1}{(k+l-1)!}(1+o_{l,k}(1))(\log \log n)^{k+l} \log \log \log n\right)$$

we get

$$\lambda_l(g(n)) = n \exp\left(-\frac{1}{(k+l-1)!}(1+o_{l,k}(1))(\log \log n)^{k+l} \log \log \log n\right).$$

As for  $\phi(g(n))$  we note that unless  $g(n) = \phi_k(n)$ ,  $g(n)$  can be written as  $\phi_l(h(n))$  where  $h(n)$  is a  $(k-l)$ -fold iterate beginning with a  $\lambda$ . From above we can see that

$$h(n) = n \exp\left(-\frac{1}{(k-l-1)!}(1+o_k(1))(\log \log n)^{k-l} \log \log \log n\right)$$

and so  $\phi(h(n))$  is bounded above by  $h(n)$  and below by

$$\begin{aligned}
\frac{h(n)}{e^\gamma \log \log h(n) + \frac{3}{\log \log h(n)}} &= \frac{h(n)}{e^\gamma \log\left(\log n - \frac{1}{(k-l-1)!}(1+o_k(1))(\log \log n)^{k-l} \log \log \log n\right)} \\
&= \frac{h(n)}{e^\gamma \log \log n - O\left(\frac{1}{(k-l-1)! \log n}(1+o_k(1))(\log \log n)^{k-l} \log \log \log n\right)} \\
&= h(n) \exp\left(O(\log \log \log n)\right)
\end{aligned}$$

which is within the error of  $h(n)$ . Hence any string of  $\phi$  will not change our estimate. Therefore if  $g(n)$  is a  $k$ -fold iteration of  $\phi$  and  $\lambda$  which is not  $\phi_k(n)$ , but which begins with  $l$  copies of  $\phi$ , then

$$g(n) = n \exp \left( - \frac{1}{(k-l-1)!} (1 + o_k(1)) (\log \log n)^{k-l} \log \log \log n \right)$$

yielding our theorem.

#### ACKNOWLEDGEMENTS

The author would like to thank Greg Martin for his assistance and guidance.

#### REFERENCES

- [1] W.D. Banks, F. Luca, F. Saidak and P. Stănică. Compositions with the Euler and Carmichael Functions, *Abh. Math. Sem. Univ. Hamburg.*, **75** (2005), 215–244.
- [2] P. Erdős, A. Granville, C. Pomerance, and C. Spiro, On the Normal Behavior of the Iterates of some Arithmetic Functions, in *Analytic number theory* (Allerton Park, IL, 1989), *Progr. Math.*, **85** Birkhäuser Boston, Boston, MA, (1990), 165–204.
- [3] P. Erdős, C. Pomerance, and E. Schmutz, Carmichael’s Lambda Function, *Acta Arith.*, **58** (1991), 363–385.
- [4] H. Halberstam and H. Richert *Sieve Methods*, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], London-New York (2004) London Mathematical Society Monographs, No.4.
- [5] J.P. Kubilius, *Probabilistic Methods in the Theory of Numbers*, Translations of Mathematical Monographs, Vol. 11, American Math. Soc., Providence (1964)
- [6] G.Martin and C. Pomerance, The Iterated Carmichael  $\lambda$ -Function and the Number of Cycles of the Power Generator, *Acta Arith.*, **118** (2005), no. 4, 305–335.
- [7] H. Montgomery and R. Vaughan, *Multiplicative Number Theory I. Classical Theory*, Cambridge University Press, New York (2007)
- [8] C. Pomerance, On the Distribution of Amicable Numbers, *J. Reine Angew. Math.*, 293/294 (1977), 217–222.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF BRITISH COLUMBIA, ROOM 121, 1984 MATHEMATICS ROAD, VANCOUVER, BC, V6T 1Z2, CANADA  
*E-mail address:* harlandn@math.ubc.ca